

# Stock price dynamics and volatility : a high frequency data perspective

## Citation for published version (APA):

Frijns, B. P. M. (2004). Stock price dynamics and volatility : a high frequency data perspective. Maastricht: Datawyse / Universitaire Pers Maastricht.

## Document status and date:

Published: 01/01/2004

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# STOCK PRICE DYNAMICS AND VOLATILITY: A HIGH FREQUENCY DATA PERSPECTIVE

CONFIDENTIAL

## STOCK PRICE DYNAMICS AND VOLATILITY: A HIGH FREQUENCY DATA PERSPECTIVE

by John C. Hull, University of Toronto, and Mark S. Pagan, University of California

Working Paper No. 95-01, University of Toronto, and University of California, San Diego, 1995

Copyright © 1995 by John C. Hull and Mark S. Pagan

200

STOCK PRICE DYNAMICS



# STOCK PRICE DYNAMICS AND VOLATILITY: A HIGH FREQUENCY DATA PERSPECTIVE

## Acknowledgements

### PROEFSCHRIFT

ter verkrijging van de graad van doctor aan  
de Universiteit Maastricht,  
op gezag van de Rector Magnificus, Prof. Mr. G.P.M.F. Mols,  
volgens het besluit van het College van Decanen, in het openbaar te verdedigen  
op 9 december 2004 om 14.00 uur

door

Bart Pieter Marie Frijns



Promotor:

Prof. dr. P.C. Schotman

Beoordelingscommissie:

Prof. dr. C.C.P. Wolff (Voorzitter)

Prof. dr. F.C. Palm

Dr. R.J. Mahieu (Erasmus Universiteit Rotterdam)

# Acknowledgements

Upon writing these acknowledgements I realize that four year have past already. Time seems to fly when doing research and definitely evolves in different time scales then the ones addressed in this thesis. Looking back over these four years I have to admit that I worked with great pleasure and enjoyed every moment as a PhD student. Although the end of these four years has been reached, I realize that I never could have finished this thesis without the involvement of some people. Here, I would like to take the opportunity to thank those people, as well as those that have been close to me during the last four years.

My first word of gratitude goes to my daily supervisor Peter Schotman. He allowed me to work independently, yet was always there when I needed him. He taught me to be self-critical and at the same time always showed his confidence in me. I am greatly indebted to him for all the effort and time he invested in me. He has always been a great source of knowledge to me and I thank him for letting me benefit from all this knowledge. I also thank him for giving me the opportunity to stay at Stockholm School of Economics, which was a very enervating experience to me.

Secondly, I would like to thank Thorsten Lehnert. One of the chapters in this thesis has been drawn from joint research with him. I have certainly benefited from all the fertile discussions that we had. Furthermore, I would like to thank Rog r Otten. He was the first to recognize a researcher in me and he greatly inspired me to become a PhD student.

I would further like to thank all my colleagues at Maastricht University for creating a pleasant working environment. In special I would like to thank all my former roommates. There have been quite a few, so without mentioning explicit names I would like to thank them all. Thanks also goes to the fellow PhD students and junior lecturers. I enjoyed the Friday afternoon "borrels", the soccer matches and the many cycling and walking trips that we made. I thank the ladies of the secretary office for their kind assistance. A special thanks goes to Rob, who was not only a colleague, but has been a very good friend for many years.

Pornpimol, I want to thank you as well. Although being far away, you recognized the importance of this for me and always motivated me to do the best I can, even if it meant putting yourself aside. You have also shown me that there are more things in life that matter, something I might forget sometimes when working.

Mijn laatste woord van dank gaat naar mijn familie. Ik dank mijn broer en zus die mij altijd, zoals ook in deze, hebben bijgestaan. Mijn grootste dank gaat echter naar mijn ouders. Zij hebben mij overal in bijgestaan, hebben altijd vertrouwen in mij gehad en stonden altijd voor mij klaar als dit nodig was. Zonder jullie was dit proefschrift er zeker niet gekomen en ik dank jullie voor alles.

Auckland, October 2004.

Acknowledgements

# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Outline	3
<b>2 Market Microstructure and the Nasdaq</b>	<b>7</b>
2.1 Introduction	7
2.2 Frictions	7
2.3 A Model for Prices	9
2.4 Price Discovery	14
2.5 The Nasdaq and its Recent Reforms	16
2.5.1 The 1997 Reform and its Impacts	18
2.5.2 Electronic Communication Networks	19
<b>3 Nonlinear Dynamics in Nasdaq Dealer Quotes</b>	<b>21</b>
3.1 Introduction	21
3.2 The Model	23
3.3 Data	27
3.4 Results	31
3.4.1 The linear model	31
3.4.2 The Nonlinear Model	33
3.4.3 Non-Linear Impulse Responses	42
3.5 Conclusion	48
<b>4 Price Discovery in Tick Time</b>	<b>49</b>
4.1 Introduction	49
4.2 A Model for Quotes in Tick Time	51
4.3 Price Discovery	54
4.3.1 Price Discovery in Tick time	54
4.3.2 Calendar Time Aggregation	56

4.4	Data . . . . .	58
4.5	Results . . . . .	61
4.5.1	Parameter Estimates . . . . .	61
4.5.2	Price Discovery in Quote Updates . . . . .	63
4.5.3	Calendar time aggregated Price Discovery . . . . .	67
4.6	Conclusion . . . . .	75
4.A	Appendix: Derivation of the Moment Conditions . . . . .	76
<b>5</b>	<b>Realized Variance in the Presence of non-IID Microstructure Noise</b>	<b>79</b>
5.1	Introduction . . . . .	79
5.2	Realized Variance . . . . .	81
5.3	Microstructure Noise . . . . .	83
5.3.1	The Effects of Microstructure Noise on Realized Variance . . . . .	84
5.3.2	A structural model for the price process . . . . .	86
5.3.3	A Model Free Correction . . . . .	89
5.4	Data . . . . .	90
5.5	Results . . . . .	92
5.5.1	Uncorrected Realized Volatility . . . . .	92
5.5.2	Structural Model: Parameter Estimates . . . . .	94
5.5.3	Structural Model: Realized Volatility . . . . .	98
5.5.4	Model-Free Correction . . . . .	100
5.6	Conclusion . . . . .	106
5.A	Appendix: Identification of the univariate state space model . . . . .	108
<b>6</b>	<b>The Information Content of Trades and Quotes</b>	<b>111</b>
6.1	Introduction . . . . .	111
6.2	Model . . . . .	114
6.3	Data . . . . .	119
6.4	Results . . . . .	121
6.5	Model Extension . . . . .	126
6.6	Conclusion . . . . .	133
<b>7</b>	<b>Summary and Concluding Remarks</b>	<b>135</b>
	<b>Nederlandse Samenvatting</b>	<b>145</b>
	<b>Curriculum Vitae</b>	<b>149</b>

# Chapter 1

## Introduction

*Among the plays which men perform in taking different parts in this magnificent world theater, the greatest comedy is played at the Exchange. There, . . . , hiding places, concealment of facts, quarrels, provocations, mockery, idle talk, violent desires, collusion, artful deceptions, betrayals, cheatings, and even tragic end are to be found.*

This excerpt from Joseph de la Vega's (1688) manuscript *Confusion de Confusiones* on the Amsterdam Stock Exchange, reveals practices still thought to be present in today's financial markets. Collusion among Nasdaq market makers led to several law suits in 1994-1995 and ultimately resulted in an alteration of the way trade was conducted at Nasdaq.<sup>1</sup> Market microstructure research is at the heart of analyzing these practices and proves to be useful for both regulators and investors.

According to Madhavan (2000) "Market microstructure studies the process by which investors' latent demands are ultimately translated into prices and volumes", whereas O'Hara (1995) defines it as "the study of the process and outcomes of exchanging assets under explicit trading rules". Ultimately, these two definitions capture the essence of market microstructure, where the translation of investors' demands can be understood as the price formation process and the different trading rules referring to different trading mechanisms. Complementary to these definitions, we argue that market microstructure studies the price formation process under alternative trading mechanisms. Understanding these mechanisms is crucial as they result in different frictions. Identifying the sources of these frictions enables regulators to adjust or introduce new trading rules, making trade more efficient. Hence, market microstructure research directs its interest towards sources, properties and consequences of frictions.

---

<sup>1</sup>The evidence for colluding market makers was provided by Christie and Schultz (1994), illustrating the importance of microstructure research in practice.

Additionally, the importance of market microstructure relates to the dynamic nature of the market for capital markets. A clear illustration of this is the evolution of exchanges through time. Where there were about 100 regional exchanges in the U.S. around 1900, this number decreased to 35 by 1935, 15 by 1965 and only 5 remain nowadays (Blume and Goldstein (1997)). The cause of this, the improvement of communication networks, makes the physical location of an exchange less important. De-regulation across Europe caused many national stock markets to merge and led to the expansion of U.S. exchanges to the European market.<sup>2</sup> Added to this are the recent developments in information technology that have facilitated the existence of many electronic exchanges and (internet based) trading venues. In essence these exchanges can be accessed from any computer terminal linked to the exchange system, greatly enhancing the liquidity of such a market. Some of these markets still require some intervention of third parties (market makers), but others do not. Alternative trading venues, might not require intervention of third parties, but are linked to an exchange that does.

These recent developments, with mainly the linking of alternative trading venues to an exchange, add to the complexity of the operations of financial markets, as the flow of information among these venues has to be ascertained. Whereas past research mainly focussed on so called real frictions in market (the costs associated with processing orders, see Stoll (2000)) current research focusses on the flow and distribution of information (informational frictions). When market participants are differentially informed the market becomes less efficient and less transparent. When the informational asymmetries are large, the market becomes a risky place, in the sense that the market price may be a poor indicator of the real value of the asset. With the increased complexity of markets, assuring a transparent and informational efficient market is essential.

The importance of market microstructure research has also reached beyond its own field of studying frictions and has found its way into other areas of finance. An example of this is in asset pricing, where the presence of informational asymmetry causes a risk that investors need to be compensated for. Easley, Hvidkjaer, and O'Hara (2002) show that stocks with a higher probability of informed trading command a risk premium and thus have a higher rate of return. They show that two similar stocks for which the probability of informed trading differs by 10 %, the difference in annual returns is 2.5%. Another area is that of volatility. Recent research allows us to infer daily volatilities from intra-day prices. These intra-day prices, however, are drastically affected by microstructure effects. It is therefore essential to properly take care of these microstructure effects (see e.g. Hansen and Lunde (2004b)).

In this study we focus on the operations conducted at the Nasdaq. The Nasdaq operates as a dealer market, where multiple market makers disseminate quotes to compete for order

---

<sup>2</sup>In 2001 the Nasdaq established itself in Europe, creating Nasdaq Europe. However, this operation failed and ceased to exist.

flow. However, after the colluding case revealed in 1994, a new Order Handling Rule was introduced. The main consequence of this rule was that public quotes can compete directly with the quotes of market makers. These public quotes can be submitted through a so called Electronic Communication Network (ECN). Trade can be conducted on these Alternative Trading Systems without ever reaching the Nasdaq and the market can become fragmented. This means that informational asymmetries can lead to temporal price differences. Moreover, if these differences occur due to information about the future value of the asset, it will have a permanent impact on prices. Knowing the size of the informational asymmetry and where the asymmetry is located is essential to analyzing the efficiency of the market. In this thesis we study these asymmetries, both among dealers and between dealers and traders.

The role of information is not only important in explaining differences among dealers/traders, but is also of great importance for the evolution of the price process. It is a well known fact in market microstructure that prices change with the flow of information. A new field in finance is developing, focussing on inferring daily volatilities from intradaily observations. Essentially, these methods try to infer the flow of information from the price process. However, the observed price process is a noisy measure of the information flow contaminated by the frictions present. This makes the flow of information unobserved. Although frictions are unobserved as well, early microstructure research has focussed on which sources of frictions exist and what properties these frictions should have. Understanding which sources of frictions exist and how these frictions affect the price process is therefore essential for understanding the flow of information. Additionally, the flow of information can be decomposed even further into a component that refers to public information and private information.

## 1.1 Outline

This section provides a short overview of the chapters in this thesis.

Chapter 2 presents a primer on market microstructure. We introduce the notion of frictions in a market. We further sketch for the price process in which these friction can be discussed. This framework is also used to discuss the empirical issues addressed in this thesis. We focus on one particular issue, called price discovery, as this forms an integral part of this thesis. As the empirical part of this thesis centers around the Nasdaq, we address some institutional characteristics of this exchange and the impact some regulatory changes have had on the operation of this market. We finally discuss some of the features of Electronic Communication Networks, which are electronic alternative trading systems, linked to the Nasdaq.

In chapter 3 we address dealer quoting behavior. We propose a nonlinear model for dealer quote interactions. Traditionally quoting behavior is addressed in a linear framework (see



Hasbrouck (1991)). This framework is based on the notion that quotes error correct towards a mid point (the average of all quotes). We argue that this correction does not occur towards the mid point but to the respective inside quotes (the best bid and ask quote in the market). Correcting further than these points makes no sense as this only leads to decreases in dealers' rents. We additionally include an indicator function that shows when a dealer is at the inside. To test for differential information among dealers we examine the reaction of each dealer to another dealer being at the inside. We finally propose to discuss price discovery in terms of impulse response functions. As the model is nonlinear we use generalized impulse response functions. These functions can be found by conditionally bootstrapping the error term of the proposed model and using them to construct price paths given relevant initial situation and relevant shocks. The results include that the sources of nonlinearity are highly significant and indicate a symmetric quote adjustment of dealers to one particular dealer. We further find an asymmetric adjustment of quotes to the size of the inside spread. The impulse response functions reveal that the inclusion of nonlinearities in the model are of great importance and lead to different results than observed with linear impulse responses. The different initial situation of the model has no influence on the outcome of the function.

Chapter 4 is devoted to price discovery. The price discovery issue exists when there are multiple venues/market makers that trade the same asset. Informational asymmetries may cause prices to be fragmented. However, in the long run prices at these different venues follow the same trend. Price discovery studies to which extent different venues contribute to the price process. The traditional measure for price discovery, referred to as the Information Share, considers the amount of variance that each venue contributes to the variance of the price process (see Hasbrouck (1995)). In practice this measure is computed using aggregated data. But, when data is aggregated different venues may already have reacted to each other's price changes. This causes prices to be contemporaneously correlated, and this correlation hinders a unique decomposition of the price discovery to each venue. In chapter 4 we propose to use a tick time model for dealer quote interactions using ultra-high-frequency data. As the model is designed for sampling at the highest frequency, the correlation issue is minimized. The empirical application is on Nasdaq market makers. To assess the price discovery we define several measures in tick time. These tick time measures can be aggregated and we define a new measure for price discovery in calendar time.

Chapter 5 addresses the issue of realized volatility. Realized volatility, introduced by Andersen, Bollerslev, Diebold, and Labys (2001), is computed by summing squared intra-day returns and provides a consistent measure for the daily volatility of an asset. The key strength of realized volatility is that it is observed rather than latent, such as the conditional volatility computed with a GARCH-type process. From a microstructure point of view, it is the flow of information that causes volatility in prices. However, at very high frequencies prices are contaminated by microstructure noise. Recent literature has proposed several approaches to

circumvent this issue by sampling at lower frequencies. The approach we follow in chapter 5 is aimed at removing the microstructure noise from the price process. We follow two methods to correct the data for microstructure noise. The first method is based on a structural model that reveals many insights in the properties of the microstructure noise. The second approach, recently introduced by Hansen and Lunde (2004b), is model-free and reveals that there are two sources affecting the price process. The first corresponds to the known microstructure effects, which biases realized volatility upwards if not accounted for. The second corresponds to a long persistence in observed returns. We question whether this persistence is a source of noise or whether this persistence is a feature inherent to the price process.

In chapter 6 we study the amounts of public and private information present in the market. Few studies have focussed on these absolute amounts<sup>3</sup>, but those who do obtain the amount of private information from the autocorrelation present in order flow (see Madhavan, Richardson, and Roomans (1997)) and is based on the following notion. In a market, the market makers set prices conditional on the public information (and their personal inventory imbalance). Traders can execute order given the prices set by the market makers. If certain traders know that the true price of the asset deviates from the price quoted by the market makers, they will execute trades at only one side of the market and thus induce autocorrelation in order flow. However, this approach either assumes that the side of the transaction is known (buy or sell) or else relies on a particular trade matching algorithm (Lee and Ready (1991)). However, for the Nasdaq, the market we focus at, the performance of this algorithm is poor. In chapter 6 we compare the information content present in trades and quotes. Comparing the information content of both reveals the amount of public and private information in both processes. Under the assumption that market makers set prices conditional on the public information in the market, quotes reveal the amount of public information. Trades, consequently, can be executed at these prices. When no private information is present, trades will occur, with equal probability, at both sides of the market and thus only generate noise around the efficient price as determined by the market makers. When private information is present trades will have an impact on the efficient price and thus trades reveal the amount of private information present. We further test for the impact of additional factors on the informativeness of both processes.

Chapter 7 summarizes the main findings of this thesis and concludes.

---

<sup>3</sup>This issue has been explored to some extent in foreign exchange market (see e.g. Lyons (2001)).



## Chapter 2

# Market Microstructure and the Nasdaq

### 2.1 Introduction

In this chapter we provide a brief introduction into market microstructure, with an emphasis on the Nasdaq. For general discussions on market microstructure theory we refer to O'Hara (1995) and Madhavan (2000), an overview on empirical issues is provided by Coughenour and Shastri (1999). We discuss several frictions that lead to costs, which market makers need to be compensated for. We further sketch a general framework in which to consider these frictions. This framework is directed towards the issues addressed in the remainder of this study. As the empirical analysis in this thesis centers around Nasdaq stocks, we elaborate on the operation of this market. Additionally, we address some issues that have affected trading at Nasdaq. One of the largest impacts on the nature of Nasdaq was the introduction of the Order Handling Rule in 1997. We address this rule and its consequences in more detail. Finally, we elaborate on Electronic Communication Networks and their operations.

### 2.2 Frictions

The field of market microstructure deviates from many other fields in finance as it recognizes the fact that prices are contaminated by several market frictions. Here we provide an overview of the frictions observed in financial markets. These frictions are addressed in more detail in Stoll (2000). Frictions lead to several costs, that a market maker must bear. As a market maker wants to be compensated for these costs, she will quote two different prices at which she wants to buy (bid quote) or sell (ask quote). The price at which she wants to sell is always higher than the price at which she wants to buy. This creates a positive difference between the two prices called the spread and represents the income she gains from a round

trip transaction (a consecutive buy and sell transaction). The spread can therefore be seen as the compensation to the market maker for frictions. The larger the spread, the more frictions present. Hence, the spread has become a measure for the magnitude of frictions.

There are several sources that cause frictions in stock markets. Knowing these sources and their magnitudes is essential for both investors and regulators. When regulators know that frictions are mainly driven by one particular source, they can alter the trading mechanism or impose rules aimed at reducing these frictions. Investors are interested in these frictions as they result in trading costs they incur. Lowering frictions reduces spreads and therefore trading costs.

According to Demsetz (1968) the spread purely reflects the costs of providing immediacy. These are the costs associated with the execution and settlement of trades and are often referred to as order processing costs. In the presence of only order processing costs, quotes are centered symmetrically around the true price and spreads remain fairly constant. The only dynamics in spreads comes from the quoted size (high quoted volume would create economies of scale in the order processing costs, so spreads would decrease in quoted volume).

A second source of frictions reflects the compensation for the unwanted inventory position that a market maker obtains when providing immediacy (see Garman (1976) and Amihud and Mendelson (1980)). A clear rationale for understanding inventory models is provided by Madhavan (2000). He notes that "physical market places consolidate buyers and sellers in *space*, the market maker can be seen as an institution to bring buyers and sellers together in *time* through the use of inventory". An unwanted inventory position leads to a risk that the market maker must be compensated for. She therefore quotes a larger spread compared to when she only faces order processing costs. In contrast to order processing costs, inventory related costs cause dynamics in the quoted prices of a market maker. For example, when the market maker receives a sell order (i.e. a transaction at the bid price), her inventory position increases. When this position is unwanted she will lower both her bid and ask quotes, making her ask price more attractive and her bid less attractive. By adjusting her bid and her ask quote she maintains a stable inventory position. Also, spreads increase in quoted volume as the receipt of more inventory leads to a greater distance from the preferred inventory position.

A less known source of frictions is referred to as market power (see Stoll (2000)). Market makers with market power will increase the spread above the level of their costs. Maintaining spreads above this level leads to higher profits for the market maker. This component explains the large observed spreads at Nasdaq and has been at the heart of a dispute about Nasdaq market makers initiated by Christie and Schultz (1994).

The last source of frictions, which are caused by asymmetric information, is referred to as informational frictions (see Stoll (2000)). This asymmetric information component was addressed by Bagehot (1971), Glosten and Milgrom (1985), Kyle (1985) and Glosten and

Harris (1988), to name but a few. Asymmetries lead to a cost component in the spread for the following two reasons. First, market makers face the risk that some traders can react more timely to the release of new information, whereas the process of setting a new quote takes some time. This reason explains the earlier existence of SOES bandits (Harris and Schultz (2002)), day traders profiting from the sluggishness of market makers on the Small Orders Execution System at Nasdaq. A second, and more known reason is the notion that there are traders with superior information. Trading with these informed traders leads to losses for the market maker. These losses must be regained from trading with the uninformed. The quoted spread reflects the compensation a market maker must obtain from trading with the informed.

In the discussion above we have introduced several cost components that lead to frictions in markets. The next section introduces a common model for prices, in which these frictions can be considered.

## 2.3 A Model for Prices

This section introduces a general framework for observed prices and illustrates how frictions can be seen within this framework. The framework is common to many studies and is used by Madhavan (2000) to integrate these studies. The most important notion of this framework is that different frictions prevent market prices from being efficient. Before we address the consequences of frictions on prices in this framework, we first define a frictionless market.

One of the most widely accepted notions in microstructure theory is the existence of a true price. Although this price is typically unobserved, it can be seen as the underlying value of an asset conditional on *all* information. However, not all information is available to all agents in the economy and we begin our discussion by restricting ourselves to the following setting. Analogous to Madhavan (2000) we assume that all agents have homogeneous information sets, the public information set ( $\mathcal{F}_t$ ). Since all agents share the same information, prices are (semi-strong) efficient in the sense that current prices reflect future price expectations conditional on the available information set. Stated differently, prices satisfy the properties of a martingale, i.e.  $E[\mu_{t+1}|\mathcal{F}_t] = \mu_t$ , where  $\mu_t$  refers to the (log) efficient price. Consequently, the efficient price follows a random walk

$$\mu_t = \mu_{t-1} + e_t, \quad (2.1)$$

where  $e_t$  is the innovation in public beliefs.

This concept of an efficient price provides a good starting point for discussing microstructure issues. The existence of frictions leads to a departure from prices being efficient, resulting in two different prices that market makers trade at. The price observed ( $p_t$ ) equals the efficient

price and a component that refers to frictions

$$p_t = \mu_t + u_t, \quad (2.2)$$

where  $u_t$  measures the size of these frictions. The observed price can be seen as the transaction price in the market in which case  $u_t$  is positive/negative for a buy/sell transaction and zero for a transaction at the mid point.

The frictions discussed in the previous section can be seen as those components that determine the size of  $u_t$ . Furthermore, these frictions all have different impacts on the (time series) properties of observed prices, returns and spreads (see Roll (1984) for a graphical illustration on the consequences of these frictions). A market with only order processing costs is characterized as follows. Under the assumption that quoted depths do not affect spreads (an assumption often adhered to), quoted prices will be symmetric around the efficient price ( $p_t^B = \mu_t - c$ ,  $p_t^A = \mu_t + c$ , where  $p_t^B$  and  $p_t^A$  refer to the bid and ask quotes respectively and  $c$  is a constant). The observed transaction prices are the realizations of these quotes and take on any of these two values. Let  $Q_t$  be a buy/sell indicator equal to 1 if the transaction was a buy for the trader and -1 if the transaction was a sell. In this case  $u_t = Q_t c$  and the observed transaction price is equal to

$$p_t = \mu_t + Q_t c. \quad (2.3)$$

When only order processing costs are present  $Q_t$  will be i.i.d., having equal probability of being 1 or -1. This causes transaction prices to bounce between the values  $\mu_t - c$  and  $\mu_t + c$ . This bounce induces a negative first order autocorrelation in observed transaction returns. As buy and sell transactions occur with equal probability the negative first order autocorrelation reaches its minimum value of -0.5 if  $\mu_t$  remains constant over time. Further, when only order processing costs are present, market makers do not adjust quotes after the occurrence of a transaction. Therefore, the quoted spread will be equal to the traded spread (the absolute difference between prices at which two consecutive buy and sell transactions are traded). Although the quoted and traded spreads are typically computed as daily averages, we here consider them per trade, assuming that  $\mu_t$  is constant overtime. Although this assumption is in fact unrealistic, it serves the purpose of the current exposition. We refer to Stoll (2000) for a more elaborate discussion. The quoted spread ( $QS_{t+1}$ ) is given by

$$QS_{t+1} = p_{t+1}^A - p_{t+1}^B = \mu_{t+1} + c - (\mu_{t+1} - c) = 2c. \quad (2.4)$$

Likewise the traded spread ( $TS_{t+1}$ ) is

$$TS_{t+1} = E[p_{t+1}|Q_{t+1} = 1] - E[p_t|Q_t = -1] = \mu_{t+1} + c - (\mu_t - c) = 2c. \quad (2.5)$$

Hence the presence of order processing costs can be identified by considering quoted and traded spreads and by the presence of negative autocorrelation in transaction returns.

Inventory costs have a different impact on the properties of  $u_t$ . In contrast to order processing costs,  $u_t$  will vary asymmetrically around the efficient price. The reason for this is that the market maker will adjust her bid and ask quotes, such that she remains a stable inventory position. For example, when a market maker receives inventory (a transaction occurring at the bid), she lowers her bid and ask quote to induce a transaction at the ask price. This results in a lower traded spread than a quoted spread. We illustrate this in the following example. We start with the same situation as for order processing costs where  $p_t^A = \mu_t + c$  and  $p_t^B = \mu_t - c$  and again assume that the efficient price does not change. When a sell transaction takes place ( $Q_t = -1$ ), i.e. a transaction at the bid price, the market maker will lower both her quotes, where  $E[p_{t+1}^A | Q_t = -1] = \mu_t$  and  $E[p_{t+1}^B | Q_t = -1] = \mu_t - 2c$ , to induce a transaction at the ask. The quoted spread in this example is easily found by subtracting the bid from the ask price,  $QS_{t+1} = p_{t+1}^A - p_{t+1}^B = 2c$ . The traded spread is found by considering two consecutive trades that occur at opposite sides of the market. Hence, given that the trade at time  $t$  was at the bid ( $Q_t = -1$ ), the traded spread is found by subtracting the traded bid price at time  $t$  from the traded ask price at time  $t + 1$ , i.e.

$$TS_{t+1} = E[p_{t+1} | Q_{t+1} = 1, Q_t = -1] - E[p_t | Q_t = -1] = c, \quad (2.6)$$

where

$$E[p_{t+1} | Q_{t+1} = 1, Q_t = -1] = E[p_{t+1}^A | Q_t = -1] = \mu_t, \quad (2.7)$$

is the traded ask price at  $t + 1$  and

$$E[p_t | Q_t = 1] = p_t^B = \mu_t - c, \quad (2.8)$$

the traded bid price at time  $t$ . The traded spread is thus half the quoted spread.

The consequences of asymmetric information on (2.2) will be explained in some more detail, as this component is of greater importance to the remainder of this thesis. The presence of asymmetric information leads to the recognition that there are two types of investors, informed and uninformed. The usual definition of an uninformed trader is a trader who has access to the public information available ( $\mathcal{F}_t$ ). Her trading is motivated by liquidity reasons. An informed trader is a trader with superior information. Based on her information about the location of the 'true' price, she will trade whenever it is profitable to her. With the traditional assumption that asymmetric information is only present among traders, this asymmetry manifests itself on one side of the market. On the one hand, market makers are those with access to public information and disseminate quotes on the basis of this information. On the other hand, traders present a mixture of informed and uninformed. An



obvious way to identify the informed from the uninformed is by comparing the trade and the quote process.

Glosten and Harris (1988) (henceforth GH) propose a model to identify the asymmetry in information. In their model uninformed traders have no permanent impact on prices as the information is already incorporated in the market (i.e. in the bid/ask spreads). The informed traders, do have a permanent impact on prices, as their trading reveals new information about the location of the 'true' price. After an informed trader has traded, the public information set ( $\mathcal{F}_t$ ) is updated to include this information. GH argue that when "asymmetric information exists, informed traders profit by submitting orders that will be correlated with future price changes". This correlation manifests itself by more trades being executed on one side of the market. The trade indicator  $Q_t$  is therefore not i.i.d. anymore. For example, when an informed trader knows that the 'true' price lies above the current price, she only trades at the ask. The uninformed traders trade at both sides of the market with equal probability. This induces autocorrelation in order flow. If there are no informational asymmetries  $E[Q_t Q_{t-1}] = 0$ . When private information is present  $E[Q_t Q_{t-1}] > 0$ . The autocorrelation is positive as private information induces more trades to occur on one side of the market.

The model of GH is based on the induced autocorrelation in order flow and is very similar to the framework presented above. They decompose the traded price into two component similar to the decomposition in (2.2),

$$p_t = \mu_t + Q_t C_t, \quad (2.9)$$

where  $C_t$  is a function of traded volume. This part reflects real frictions, such as order processing costs and inventory costs, and as such does not have a permanent impact on the price of an asset. Second they model the evolution of the efficient price as

$$\mu_t = \mu_{t-1} + e_t + Q_t Z_t, \quad (2.10)$$

where  $Z_t$  measures the informational asymmetry, again a function of volume. The first part of the equation is identical to (2.1) and represents the evolution of the efficient price based on public information. The second part relates to the private information. As traders with private information know the location of the true price, their trades have a permanent impact on the efficient price.

The two components referring to real frictions and information asymmetry can be estimated from the first difference in transaction prices. For simplicity we assume that volumes have no impact on  $C_t$  and  $Z_t$ , so  $C_t = c$  and  $Z_t = z$ . The first differences are given by

$$\begin{aligned} \Delta p_t &= \mu_t - \mu_{t-1} + Q_t c - Q_{t-1} c \\ &= e_t + Q_t z + Q_t c - Q_{t-1} c \end{aligned} \quad (2.11)$$

and so we can estimate the equation

$$r_t = \alpha Q_t + \beta Q_{t-1} + e_t, \quad (2.12)$$

where  $r_t = \Delta p_t$ ,  $\alpha = z + c$  and  $\beta = c$ .

An alternative approach, the one we opt for in this study, is to allow for correlation between the innovation in the efficient price ( $e_t$ ) and the component referring to frictions ( $u_t$ ). In this case  $u_t$  measures all frictions, real and informational, and allows a part of these frictions to have a permanent impact on the innovation in the efficient price. The private information reveals itself through this correlation as a transaction based on private information has a permanent impact on the innovation in the efficient price. Through this correlation we maintain the random walk property of the efficient price, but allow some market participants to be informed about its innovation.

An additional interesting feature of the GH model are the properties of the innovation in the random walk. GH assume that the innovation in the random walk is i.i.d. normal conditional on the time between transactions. Reason for this are the results of Clark (1973) and Harris (1987) who show that real time is not the appropriate time scale for the evolution of the price process. They argue that there may be another process that is driving the time evolution of the price process, such as traded volume or the number of transactions.

The framework presented above will be used throughout this thesis. In particular, in chapter 6, we use a similar motivation as Glosten and Harris (1988) to infer the amount of public and private information in the market. The idea behind the GH model is that conditional on quoted prices, investors trade. Informed traders know more about the location of the efficient price and thus trade more on one side of the market. This induces positive autocorrelation in order flows. The approach we follow is slightly different. We decompose the innovations in the efficient price as coming from the quote process and the trade process. All public information will be transmitted through the quotes set in the market. Additional information that flows from trades stems from private information. This decomposition also enables us to test for some extra variables that might affect the amount of private and public information in the market.

Another issue addressed in chapter 5, is the notion that prices change with the flow of information. However, the observed price process is a noisy measure for the flow of information, as investors who trade for liquidity purposes, do not reveal any information and trade at both sides of the market. In chapter 5 we decompose the noise component from the efficient price process and integrate the innovations in the efficient price up to a daily level. The daily innovation in the efficient price has become known as Realized Volatility.

## 2.4 Price Discovery

Another field in microstructure focusses on the flow of information from different sources. With the general notion that prices change with the flow of information, relevant questions are: how does this information flow into the market, who contributes most to the information flow and how quickly is the information incorporated into the market? The topic that covers these issues has become known as price discovery, which forms an integral part in this thesis. In this section we will address a commonly used measure for price discovery.

Consider two different venues that trade the same asset (extension to more venues are straightforward). The observed prices are  $p_{1,t}$  and  $p_{2,t}$  and share one common trend, the efficient price of the asset. This efficient price is assumed to follow a random walk. We can now extend (2.2) to a multivariate setting,

$$p_t = \iota \mu_t + u_t, \quad (2.13)$$

where

$$p_t = \begin{pmatrix} p_{1,t} \\ p_{2,t} \end{pmatrix}, u_t = \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix},$$

and  $\iota$  is a unit vector. We further assume that  $u_t$  is covariance stationary. Equation (2.13) can be seen as the integrated process of random walk innovations plus the transitory deviation observed at time  $t$ . Recursive substitution of (2.1) in (2.13) yields

$$p_t = p_0 + \iota \sum_{i=0}^t e_i + u_t. \quad (2.14)$$

Likewise (2.14) can be stated in first differences,

$$\Delta p_t = \iota e_t + \Delta u_t, \quad (2.15)$$

a covariance stationary process. As all covariance stationary process can be written in their Wold representation, we can also write

$$\Delta p_t = \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \dots = \Psi(L) \varepsilon_t, \quad (2.16)$$

where  $\Psi(L)$  matrix polynomial in the lag operator. According to Beveridge and Nelson (1981),  $\Psi(L)$  can be decomposed as

$$\Psi(L) \equiv \Psi(1) + (1 - L)\Psi^*(L), \quad (2.17)$$

where

$$\Psi(1) = I + \Psi_1 + \Psi_2 + \dots,$$

the sum of all the moving average coefficients.

To arrive at a measure for price discovery we use the fact that both prices follow the same common trend. This means that both prices are cointegrated, i.e. there is a linear combination of  $p_{1,t}$  and  $p_{2,t}$  that is covariance stationary. To illustrate this, consider a zero-net investment strategy in both markets, where a security is purchased in one market and sold in the other. This investment will only lead to an exposure to the transitory component of prices in both venues,

$$(1 \ -1) \begin{pmatrix} p_{1,t} \\ p_{2,t} \end{pmatrix} = (1 \ -1) \iota \mu_t + (1 \ -1) \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} = u_{1,t} - u_{2,t},$$

which is a covariance stationary process. The vector  $\beta' = (1 \ -1)$  is called the cointegrating vector.

The fact that both prices are cointegrated implies that  $\beta' \Psi(1) = 0$ . Second, the fact that both prices are equally exposed to the efficient price ( $\beta' = (1 \ -1)$ ) and that the fact that  $\beta' \Psi(1) = 0$  means that the rows in  $\Psi(1)$  are identical. Let  $\psi$  refer to one of these rows, then substituting this into (2.16) and using the fact that all rows in  $\Psi(1)$  are equal, we obtain

$$\Delta p_t = \iota \psi \varepsilon_t + \Psi^*(L) \Delta \varepsilon_t, \quad (2.18)$$

or stated in levels as

$$p_t = p_0 + \iota \psi \sum_{i=0}^t \varepsilon_i + \Psi^*(L) \varepsilon_t. \quad (2.19)$$

So the vector  $\psi$  measures the permanent impact that prices have of the efficient price, and  $\Psi^*(L)$  measures transitory effects. Comparing (2.14) and (2.15) with (2.18) and (2.19), we see that the innovation in the efficient price ( $e_t$ ) is given by  $\psi \varepsilon_t$  and so the variance of the efficient price is given by  $\psi \Sigma \psi'$ , where  $\Sigma = \text{Var}(\varepsilon_t)$ . The transitory component ( $u_t$ ) is given by  $\Psi^*(L) \varepsilon_t$ .

Hasbrouck (1995) uses this decomposition to arrive at a measure for price discovery, referred to as the information share. The information share is defined as the amount of variance each venue contributes to the total variance of the random walk. For example, if we assume that  $\Sigma$  is diagonal, the information share of venue  $i$  is given by

$$IS_i = \frac{\psi_i^2 \Sigma_{ii}}{\psi \Sigma \psi'}. \quad (2.20)$$

However, in most cases  $\Sigma$  is not diagonal, so a unique decomposition does not exist. If

price changes in two venues are correlated one cannot determine whether the contribution to the efficient price comes from one venue or the other. The main factor driving this correlation is time aggregation. When data is aggregated over longer intervals, information will have revealed itself in both venues and prices are adjusted accordingly. In the presence of contemporaneous correlation Hasbrouck (1995) proposes alternative Choleski decompositions, to establish upper and lower bounds for the information shares.

In obtaining the parameters in  $\psi$  Hasbrouck (1995) uses the fact that a cointegrated system can be expressed as an error correction model (Granger Representation Theorem), where the parameters of the error correction model can be used to find the elements in  $\psi$  (see Hasbrouck (1995)). The error correction model takes the form

$$\Delta p_t = \sum_{j=1}^k \phi_k \Delta p_{t-k} + \alpha \beta' p_{t-1} + \eta_t. \quad (2.21)$$

Recalling that  $\beta$  is the cointegrating vector,  $\alpha$  measures the extend to which prices *correct* towards each other and  $\phi_k$  measures the impact of past price changes on each other.

In the next chapter, we extend this model to include several nonlinearities. Essentially the error correction of prices towards each other is the same as a correction towards the average price. For dealers at Nasdaq we argue that this correction is not towards the mid quote (the average of all dealer quotes), but towards the inside quotes (the best bid and offer in the market). Correcting beyond these quotes only leads to a decrease in profits and is not preferable. Some additional nonlinear factors reveal that the identity of the dealer at the inside also matters. We finally, discuss price discovery using impulse response functions, where we give a structural interpretation to the error term in the model (instead of applying alternative Choleski decompositions), to arrive at an interpretable decomposition. The impulse response functions are informative about the speed at which information is incorporated in prices.

The previous sections have focussed on microstructure models and topics we address in this thesis. The empirical research focusses on the Nasdaq, which we elaborate on in the next section.

## 2.5 The Nasdaq and its Recent Reforms

In this study we employ Nasdaq data from 1999. The following discussion focusses on the most relevant developments up to that date. The most important event that influenced the nature of how trading was conducted at Nasdaq was the introduction of the Order Handling Rule (henceforth OHR), introduced in January 1997. In this section we address these new rules and the consequences they had. However, we start by briefly describing the structure

of the Nasdaq and its key features.<sup>1</sup>

The Nasdaq (National Association of Securities Dealers Automated Quotations system) was founded in 1971. As a fully owned subsidiary of the NASD (National Association of Securities Dealers) it was originally a means for associated dealers to exchange information (quotes) for Over The Counter (OTC) traded shares. In fact it is still registered as a "securities information processor", rather than a stock exchange. In the early days it represented no more than a network of computer terminals on which dealers could enter quotes and traders could see them. The actual execution was done by contacting a specific market maker and negotiating a price. Later, when the importance of this market grew, it developed into a market where securities were actually listed and traded. As there are multiple market makers that disseminate quotes, the market is a quote driven market.<sup>2</sup> This means that market makers issue firm quotes at which they have the obligation to trade (they have to trade at this price or a better price). The exchange is electronic meaning that there is no physical location where traders come together. Traders typically observe the best quotes available (also referred to as the inside quotes or Level I quotes) and can trade against these quotes with the appropriate volume. To attract order flow from traders, market makers compete with each other to set the best price in the market. Through this competition the inside spread should reduce to reflect only the true costs involved with market making.

The description of the Nasdaq above illustrates the ideal situation of how a dealer market should operate. However, as the quote in the first chapter already indicates, the exchange is common ground for deceptions and collusion. The same holds for the Nasdaq. The competitive nature of the Nasdaq was first questioned by Christie and Schultz (1994) who found that market makers avoided setting odd-eight quotes in the market (and so avoided odd-eight spreads as well). Their conclusion that market makers tacitly collude and maintain spreads far above the level of the pure costs involved with market making, led to several investigations of the Securities and Exchange Commission (SEC) and the US Department of Justice.<sup>3</sup> Their study resulted in several follow up studies that highlighted the market making practices at Nasdaq.

As an alternative to the tacit collusion hypothesis of Christie and Schultz (1994), Godek (1996) proposes preference trading as an explanation for the odd-eight avoidance. Preference trading, which is common practise on the Nasdaq, is an agreement between market makers and brokerage firms. The brokerage firm agrees to route all its orders to a specific market maker, who in turn executes this order at the best quoted price in the market. Often the brokerage firm receives a small rebate for her order, which became known as payment for order

<sup>1</sup>For a review on the history and operation of Nasdaq we refer to Smith, Selway III, and McCormick (1998).

<sup>2</sup>This feature contrasts the NYSE (New York Stock Exchange), which is an order driven market.

<sup>3</sup>The Department of Justice actually revealed that market makers use their quotes as a signalling device. When market makers issued quotes with a spread larger than \$3/4, this informed other market makers to avoid odd-eight quotes (see Smith, Selway III, and McCormick (1998)).

flow (see e.g. Kandel and Marx (1997)). Moreover, these preference agreements became the main business for Nasdaq market makers. When a market maker has preference agreements with brokerage firms to execute at the best price in the market, she no longer has the incentive to improve the best quotes in the market. Even more, improving the quotes would only reduce the profits that she makes. Obviously, these practices hinder the competitive nature of the Nasdaq, which ultimately led to a change in trading rules.

### **2.5.1 The 1997 Reform and its Impacts**

The Order Handling Rule (OHR), introduced in January 1997, has had four major components that have implications for trading at Nasdaq. We will briefly document these four components (see Barclay, Christie, Harris, Kandel, and Schultz (1999) for a more elaborate discussion).

The first component, the Limit Order Display Rule, states that when a trader contacts a market maker and wants to trade at a better price than stated in the market, the market maker either has to execute this trade against her own inventory or display this quote as her own on the Nasdaq or route this quote to another market maker or trading system that matches this quote. This new rule allows the public to compete directly with the quotes set by the market makers, which before the change was not possible.

The second component greatly enhanced the transparency of the market. It stated that market makers cannot post better quotes at electronic communication networks (ECNs), unless this ECN reports its best quotes to the Nasdaq. Before the introduction of this rule it was common practise for market makers to post better prices on Instinet (a proprietary trading system used among market makers and institutions) than on the Nasdaq, creating a difference between wholesale prices and retail prices. This rule enhances competition and reduces the fragmentation across different trading venues.

A third rule reduced the minimum quote size for market makers. Traditionally shares were traded with a minimum of 1,000 shares (or in some cases 500). This quote size was reduced to a minimum of 100. The motivation behind this was that if the public was allowed to compete with market maker quotes, trade size had to be reduced to let the public trade and the market makers compete with the public.

The final rule is the excess spread rule (ESR). Before the OHR the rule stated that all dealers had to remain their spreads within 125 percent of an average spread. Restricting all market makers to keep their spreads within this range could motivate market makers to maintain wider spreads. Under the new regulation this rule was relaxed substantially. However it was proven that the relaxation of this rule had only little impact on trading costs.

Together with the new Order Handling Rule of the SEC, the NASD also decided to reduce the tick size. Before the introduction of the rule quoting was done in eighths of dollars. With

the introduction of the OHR this was reduced to 16<sup>th</sup>'s of a dollar (for shares with price below \$10 quoting was done in 32<sup>th</sup>'s).<sup>4</sup> Trades, however, can be executed at any price, but are rounded by the Nasdaq Trade Dissemination Service to the nearest \$1/64.

Barclay, Christie, Harris, Kandel, and Schultz (1999) were the first to address the impact of these rules on the trading costs and depths of Nasdaq stocks. They found that after the implementation of these rules spreads dropped substantially to about two thirds of the level before the implementation. Weston (2000) confirms these findings and notes that the decrease in the spread is mainly due to a decrease in the market makers' rents. He additionally finds that spreads have gone down to a level comparable to spreads quoted on the New York Stock Exchange, with which there a large discrepancy before the implementation (see e.g. Huang and Stoll (1996)).

### 2.5.2 Electronic Communication Networks

One of the main consequences of the OHR was the liberalization of ECNs. Allowing the public to compete directly with Nasdaq market makers resulted in ECNs which operate like limit order books (i.e. investors can enter quotes in an ECN). The top of the limit order book (the best bid and ask quotes) is disseminated to Nasdaq as a regular dealer quote. As such public traders can undercut the quotes issued by traditional market makers.

The growth of ECNs started almost immediately after the introduction of the OHR and is driven by the advantages of these systems offer. Firstly, trades can often be executed at better prices than quoted in the market (the main ECNs (Island and Instinet) have a minimum tick size of \$1/256, with low priced stocks having even smaller tick sizes.<sup>5</sup>). The quotes disseminated to Nasdaq are rounded outwards (up for the ask, down for the bid) to the nearest Nasdaq tick size. Hence the inside spread quoted at ECNs is often smaller than the inside spread at Nasdaq. Second, traders do not need to go to traditional brokers/market makers to have their orders executed, leading to lower trading costs. Moreover, a trade on an ECN can be executed anonymously, as the ECN takes the position of both buyer and seller in a single transaction (so it is the ECN that is reported to the Nasdaq system as the trading party).

ECNs have the same status as the traditional market makers and are registered as broker/dealer disseminating the top of their limit order book as dealer quotes. They have to comply to all the regulations as set out for market makers, with the exception that they do not have to make the market on both sides of the market continuously. This relaxation was introduced as, for illiquid stocks, the limit order book might dry up on one side of the market.

From the introduction of the OHR onwards, there have been two dominant ECNs in terms

<sup>4</sup>From April 2001 onwards quoting is done in decimals, with a minimum tick size of \$0.01.

<sup>5</sup>After the decimalization the minimum tick size at ECNs was also reduced to \$0.001.



of trading volume. Instinet ECNs, owned by Reuters plc., was the traditional ECNs where market makers and institutions traded at more favorable (wholesale) prices than the prices quoted at the Nasdaq. Nowadays it still directs its business to the larger institutions in the market. The Island ECN has found its business in the smaller public investors. It is also known that many day traders use Island as it guarantees anonymity and fast execution. Many of the other ECNs that were introduced at a later stage have failed to capture a significant market share and many of these have merged or been taken over by other ECNs or brokerage firms.<sup>6</sup>

A consequence of the rapid growth of ECNs was that a substantial amount of trading volume moved from traditional Nasdaq market makers to ECNs. Public investors trading on an ECN do not experience any inventory costs (they enter the order because they want to acquire a certain position) and because public investors only suffer from maintaining large spreads, competition increased substantially. A result of this has been that some ECNs, such as Instinet and Island have filed for exchange status with the SEC, a status which up till now has not been granted. Nasdaq's response to this increased competition with ECNs was the introduction of the SuperMontage system, which basically adopts the same structure as an ECNs but is owned by the Nasdaq. The introduction of the SuperMontage system has been a turnaround for the growth in ECNs, and many ECNs are expected to disappear in the near future.

---

<sup>6</sup>In fact Insitinet acquired the Island ECN in 2002. Although both ECNs continue to operate individually, they have been pooled in a single entity called INET.

## Chapter 3

# Nonlinear Dynamics in Nasdaq Dealer Quotes<sup>1</sup>

### 3.1 Introduction

The empirical microstructure literature on the dynamics of bid and ask quotes is abundant. Understanding these dynamics is essential for understanding how prices are formed and how markets operate. There are several factors that influence these dynamics, which can be grouped into two main categories, external and internal factors. External factors are those that occur outside the quote process, such as the impact traded volume, trade direction and stock price volatility have on quotes (see e.g. Engle and Patton (2004) and Hasbrouck (1999)). Internal factors are those that occur inside the quote process. These are, for example, a dealer's inventory position or, in the case of a multiple dealer market, informational asymmetries among dealers. Informational asymmetry is the main topic of this chapter.

Informational asymmetries across different trading venues have been addressed by Hasbrouck (1995) among others. His model measures the extent to which different venues error correct towards an (unobserved) efficient price. When this error correction is quick there will be less asymmetry, when the error correction is slow there will be more. As the efficient price is unobserved, Hasbrouck (1995) proposes to use the mid point (average of all prices) to which all prices correct (this correction was proposed by Hasbrouck (1991)). At Nasdaq, Huang (2002) applies the same methodology to dealer's bid and ask quotes to discuss the efficiency of the quote setting of these dealers.

In this chapter we argue that dealers at Nasdaq do not error correct to the midquote, but to the inside quotes in the market (i.e. the best bid and ask quote in the market). We therefore propose a model where dealer quotes correct towards the respective inside quote,

---

<sup>1</sup>This Chapter is based on B. Frijns and P. Schotman, 'Nonlinear Dynamics in Nasdaq Dealer Quotes', LIFE Working Paper WP03-002.

instead of the midquote. The model is nonlinear and generalizes the linear model used so far (see e.g. Hasbrouck (1995) and Huang (2002)). We additionally include dummy variables that indicate when a specific dealer is at the inside. These dummies should reveal whether being at the inside is informative to other dealers. We discuss price discovery in terms of dealer quoting behavior and use nonlinear impulse response functions to analyze the impact of quote changes on other dealer quotes. In our empirical analysis we consider the top two Electronic Communication Networks (ECNs), Island and Instinet, and the top three market makers, in terms of quoting frequency, of 20 actively traded stocks at Nasdaq.

Our results include that the nonlinear specification significantly improves the linear specification of Hasbrouck (1991), where the largest improvement is caused by the inclusion of the inside dummies. Dealers correct strongly towards the inside quote, but, except for Island, tend to move away from it when they have reached it. Overall, when Island sets the inside quote, dealers adjust their quotes to the inside as well. This effect is not present when any of the other dealers is at the inside. We further find that all dealers have the tendency to keep spreads small, but that this effect is strongest for ECNs. Price discovery is addressed in terms of impulse response functions. These functions not only allow us to determine the impact that a dealer's quoting behavior has on the quotes of other dealers, but also allow us to evaluate the speed at which this adjustment takes place. We compute impulse response functions for both the linear and the nonlinear model. We find that the impulse responses for the nonlinear model lead to different results than for the linear model.

The nonlinear model is motivated by the developments that have taken place at the Nasdaq. Since the studies of Christie and Schultz (1994) and Christie, Harris, and Schultz (1994) there has been a fierce discussion about competition at Nasdaq. In 1997, the Order Handling Rule (OHR), stating that all trades have to be executed against the National Best Bid or Offer (NBBO), tick size reduction and the further liberalization of ECNs (Electronic Communication Networks), have brought a great change to the competitive nature of trade at Nasdaq. Nowadays over 60% of quotes are issued by ECNs, which are also considered to be most efficient in terms of quote setting (see e.g. Huang (2002)). Efficient ECN quotes would prohibit potential collusion at Nasdaq, since ECNs would only suffer from keeping prices above the level of marginal costs. As ECN quotes are at the inside most of the times, dealer profits will reduce. Clearly, no dealer profits from setting quotes inside the inside quote. For this reason we propose an error correction towards the inside quote, instead of the midpoint (see e.g. Hasbrouck (1991)).

Since ECN quotes are at the inside most of the times, dealer quotes will become more indicative. Dealer quotes would then be a means of disseminating information through the market. When a dealer, who is normally not at the inside, sets the inside quote, this can signal information about either her inventory position or information she might possess. We therefore argue that being at the inside reveals information and other dealers might respond

to this and include this information as a second source of nonlinearity.

The remainder of this chapter is structured as follows. In the next section we propose a nonlinear adjustment model for dealer quotes, which is a generalization of the linear model. We further show how, given the decomposition proposed above, dealer efficiency can be measured with nonlinear impulse-responses. In section 3.3 we address the data used in this study. Section 3.4 presents the empirical results of our models and discusses the results of the impulse responses. Finally section 3.5 concludes.

## 3.2 The Model

In this section we propose a model for the dynamic interaction between dealer quotes. We show that this model can be seen as a generalization of the linear model proposed by Hasbrouck (1991). We further propose an alternative decomposition for quotes, which we can analyze in an impulse response framework.

For an actively traded stock at Nasdaq we consider bid and ask quotes from the  $N$  most active dealers in terms of quoting frequency. For each dealer  $i$ , where  $i = 1, \dots, N$ , let  $b_{i,t} \equiv \log(\text{Bid}_{i,t})$ , where  $\text{Bid}_{i,t}$  is the bid quote of dealer  $i$  at time  $t$ . Likewise define  $a_{i,t} \equiv \log(\text{Ask}_{i,t})$ . These quotes are stacked in the vectors  $b_t \equiv (b_{1,t}, \dots, b_{N,t})'$  and  $a_t \equiv (a_{1,t}, \dots, a_{N,t})'$ . We further define the inside quotes  $biq_t$  as the highest bid over all dealers and  $aiq_t$  as the lowest ask over all dealers. These inside quotes are determined over all dealer quotes and not only the  $N$  most active. Therefore the inside quotes enter the model exogenously. We propose the following nonlinear vector error correction model for the dynamic interaction in dealer quotes,

$$\begin{aligned} \begin{bmatrix} \Delta b_t \\ \Delta a_t \end{bmatrix} &= D_t(\cdot) + \Phi \begin{bmatrix} \Delta b_{t-1} \\ \Delta a_{t-1} \end{bmatrix} + \Psi \begin{bmatrix} (b_{t-1} - \iota biq_{t-1}) \\ (a_{t-1} - \iota aiq_{t-1}) \end{bmatrix} \\ &\quad + \theta(aiq_{t-1} - biq_{t-1}) + \begin{bmatrix} \eta_t \\ \nu_t \end{bmatrix}, \end{aligned} \quad (3.1)$$

where  $\Phi$  captures the AR(1) component<sup>2</sup>,  $\Psi$  contains coefficients on the nonlinear error correction component,  $\theta$  considers the exposure to the inside spread and  $\iota$  is a  $(N \times 1)$  unit vector.  $D_t(\cdot)$  is a deterministic function of constants and dummies, including a dummy for the overnight return, but is not considered for the moment. The innovation terms  $\eta_t$  and  $\nu_t$  are assumed to independently distributed, and we will later introduce a more specific structure on their covariances.

<sup>2</sup>Of course extensions to longer lags are straightforward.

To see (3.1) as a generalization of the linear model, partition  $\Psi$  and  $\theta$  as

$$\Psi = \begin{pmatrix} \Psi_{bb} & \Psi_{ba} \\ \Psi_{ab} & \Psi_{aa} \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_a \\ \theta_b \end{pmatrix}. \quad (3.2)$$

Equation (3.1) reduces to a linear VECM if

$$\begin{pmatrix} \Psi_{bb} & \Psi_{ba} \\ \Psi_{ab} & \Psi_{aa} \end{pmatrix} \begin{pmatrix} -\iota & 0 \\ 0 & -\iota \end{pmatrix} \begin{pmatrix} biq_{t-1} \\ aiq_{t-1} \end{pmatrix} + \begin{pmatrix} \theta_b \\ \theta_a \end{pmatrix} \begin{pmatrix} -1 & 1 \end{pmatrix} \begin{pmatrix} biq_{t-1} \\ aiq_{t-1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (3.3)$$

where  $\iota$  is a  $(N \times 1)$  vector of ones. This results in the following  $4N$  parameter restrictions

$$\begin{aligned} (\Psi_{bb} + \Psi_{ba})\iota &= 0, & \theta_b &= -\Psi_{bb}\iota, \\ (\Psi_{ab} + \Psi_{aa})\iota &= 0, & \theta_a &= \Psi_{aa}\iota. \end{aligned} \quad (3.4)$$

In the restricted model the rows of  $\Psi$  add up to zero, which imposes linear cointegration upon the model. With these restrictions the inside quotes disappear from the model. A Wald test, considering a  $\chi^2$  with  $4N$  degrees of freedom, can be used to test the linear model against the nonlinear one.

In addition to the inclusion of inside quotes we extend the linear model with inside dummies ( $D_t^I$ ) in the deterministic function  $D_t(\cdot)$ . These dummies are equal to one when a particular dealer reaches the inside quote and zero otherwise. The inclusion of these dummies is motivated by the fact that being at the inside might reveal information. For example, when dealer  $i$ , who is seldom at the inside, reaches the inside, information might be revealed to other dealers. The opposite also holds: a dealer who often quotes the inside, might reveal information when not quoting the inside (see e.g. Stoll (1989)).

As being at the inside might alter the behavior of a dealer (quoting with specific information can be different from quoting without specific information) we also expect regime dependent heteroskedasticity. We define  $\Omega_t \equiv [(\eta_t' \nu_t')' (\eta_t' \nu_t)']$  as the conditional covariance matrix of quote innovations at time  $t$ . To test for regime dependent heteroskedasticity we perform a Breusch-Pagan test (see Breusch and Pagan (1979)). For the dealers' bid and ask variances this is done on the regressions,

$$\Omega_{ii,t} = \gamma_{0,i} + \gamma_{1,i} D_{b,i,t}^I + \gamma_{2,i} D_{a,i,t}^I, \quad (3.5)$$

where  $\Omega_{ii,t}$  is the bid or ask variance of dealer  $i$ ,  $D_{b,i,t}^I$  is the inside dummy equal to 1 if the bid of dealer  $i$  is at the inside and  $D_{a,i,t}^I$  the inside dummy equal to 1 if the ask of dealer  $i$  is at the inside. The null hypothesis of this test is whether the parameters  $\gamma_{1,i}$  and  $\gamma_{2,i}$  are jointly equal to zero (i.e.  $H_0 : \gamma_{1,i} = \gamma_{2,i} = 0$ ). For the covariances (3.5) is extended by

including the inside dummies for the other dealer, i.e.

$$\Omega_{ij,t} = \lambda_{0,i,j} + \lambda_{1,i,j} D_{b,i,t}^I + \lambda_{2,i,j} D_{a,i,t}^I + \lambda_{3,i,j} D_{b,j,t}^I + \lambda_{4,i,j} D_{a,j,t}^I, \quad (3.6)$$

where  $\Omega_{ij,t}$  is the covariance between a particular quote of dealer  $i$  and dealer  $j$ . The null hypothesis of this test is again whether the parameters are jointly equal to zero, i.e.  $H_0 : \lambda_{1,i,j} = \lambda_{2,i,j} = \lambda_{3,i,j} = \lambda_{4,i,j} = 0$ .

To discuss dealer efficiency and price discovery we opt for an alternative approach than the approach proposed by Hasbrouck (1995). Hasbrouck (1995) considers the total amount of variance each dealer contributes to the total variance of the price process. However, a unique decomposition does not exist when quote innovations are contemporaneously correlated. This contemporaneous correlation is mainly driven by data aggregation, as over a long enough period each dealer will have innovated her quotes in the appropriate direction. We therefore opt for an alternative approach by computing impulse response functions. These functions are computed for both the nonlinear model in (3.1) and the linear specification where the restrictions in (3.4) are imposed, where the linear impulse responses are merely used to set a benchmark for the nonlinear functions.

To apply an appropriate shock to a dealer's quote, we have to specify what an appropriate plausible shock is. For this we propose to decompose the quote innovations  $\eta_t$  and  $\nu_t$  into two components. The first component refer to a market-wide shock common to all dealers and can be seen as the public information shared by these dealers. The second component is dealer specific and can be seen as the private information of a dealer. As the private information is known to only this dealer, we allow for correlations between the bid and the ask of this dealer. The impulse response functions for the linear model do not depend on the size of the shock and the history of the model, but we need the correlation between the bid and ask for each dealer. For the nonlinear model the size of the shock and the history of the model do matter (see Koop, Pesaran, and Potter (1996)).

We apply the following decomposition to the innovation terms  $\eta_t$  and  $\nu_t$ ,

$$\begin{aligned} \eta_t &= \varepsilon_t + \xi_{b,t}, \\ \nu_t &= \varepsilon_t + \xi_{a,t}, \end{aligned} \quad (3.7)$$

where  $\varepsilon_t$  represents the market or fundamental noise and  $\xi_{b,t}$  and  $\xi_{a,t}$  represent the dealers' idiosyncratic noise. We impose the following structure on the unconditional covariance matrix,

$$\begin{aligned} E[\varepsilon_t^2] &= \sigma_\varepsilon^2, \\ E[\xi_{b,t} \xi_{b,t}'] &= \Xi_{bb}, \\ E[\xi_{a,t} \xi_{a,t}'] &= \Xi_{aa}, \\ E[\xi_{b,t} \xi_{a,t}'] &= \Xi_{ba}, \end{aligned}$$

where the  $(N \times N)$  matrices  $\Xi_{aa}$ ,  $\Xi_{bb}$  and  $\Xi_{ba}$  are all diagonal. All other moments are zero. Given these assumption on the error term, the unconditional restricted covariance matrix will have the structure

$$\tilde{\Omega} = \begin{bmatrix} \sigma_{\varepsilon}^2 l' l + \Xi_{bb} & \sigma_{\varepsilon}^2 l' l + \Xi_{ba} \\ \sigma_{\varepsilon}^2 l' l + \Xi_{ab} & \sigma_{\varepsilon}^2 l' l + \Xi_{aa} \end{bmatrix}. \quad (3.8)$$

The elements in  $\Xi_{aa}$ ,  $\Xi_{bb}$  and  $\Xi_{ba}$  provide us with plausible sizes of dealer specific shocks.<sup>3</sup>

With a plausible size for the shock we can determine the nonlinear impulse responses. For this we follow the approach of Koop, Pesaran, and Potter (1996), who consider Generalized Impulse Response Functions (*GI*). These *GI* are dependent on the history of the model ( $y_{t-1} = (a_{t-1}, b_{t-1})$ ), and the size and direction of the shock ( $v_t$ ). The Generalized Impulse Response Function (*GI*) is defined as

$$GI(k, v_t, y_{t-1}) = E[y_{t+k}|v_t, y_{t-1}] - E[y_{t+k}|y_{t-1}], \text{ for } k = 0, 1, \dots, K \quad (3.9)$$

where  $k$  is the number of periods considered. The conditional expectation  $E[y_{t+k}|v_t, y_{t-1}]$  considers the expected path of quotes conditional on the shock that has been applied and conditional on a specific history. The conditional expectation  $E[y_{t+k}|y_{t-1}]$  leads to the expected path conditional on the specific history. This is also called the baseline. To arrive at unconditional nonlinear impulse responses we can integrate over all possible shocks and all possible histories, however for the present setting we are interested in specific shocks and histories.

To construct the baseline for these *GI* we integrate out all possible paths given a specific history. For this baseline we thus start given a specific initial situation, e.g. a specific dealer at the inside, and then simulate the possible paths from this initial situation. One could use a bootstrap to construct this baseline, drawing from the error terms of the model, but this approach does not work for (3.1). The reason for this is that we have to maintain positive spreads at all times. When randomly drawing from the error term this is not guaranteed. Hence we opt for an alternative approach. We simulate the system  $(K + 1)$  steps ahead by conditionally bootstrapping the innovation terms from our estimated model. We call the bootstrap conditional since we draw from the innovation term distribution in such a way that spreads remain positive.<sup>4</sup> A second reason for doing a conditional bootstrap is the regime dependent heteroskedasticity that might be present in the innovation terms. The distribution of the error term could be different depending on whether dealer  $i$  would be at the inside or

<sup>3</sup>Our main interest is in the dealer specific shocks as they indicate how other dealers react to private information.

<sup>4</sup>To perform this conditional bootstrap correctly we would have to draw from the error term of our model conditional on the state the model is in. Or stated differently, if dealer  $i$  is at the inside, we have to find the innovation terms where this dealer  $i$  is at the inside and draw from this distribution. However, since we are modelling the behavior of 5 dealers, we would have 10! possible distributions to draw from. Since this number is much larger than the number of observations that we have, this is empirically not feasible.

Table 3.1: List of Company Ticker symbols and Company names

Ticker Symbol	Company name
AAPL	Apple Computer Inc.
AMAT	Applied Materials Inc.
AMGN	Amgen Inc.
AMZN	Amazon.com Inc.
ATHM	At Home Corporation
CMGI	CMGI, Inc.
COMS	3Com Corporation
CPWR	Compuware Corporation
CSCO	Cisco Systems Inc.
DELL	Dell Computer Corporation
EGRP	E*TRADE Group, Inc.
INTC	Intel Corporation
MSFT	Microsoft Corporation
NOVL	Novell Inc.
NXTL	Nextel Communications CL-A
ORCL	Oracle Corporation
PSFT	Peoplesoft Inc.
SUNW	Sun Microsystems Inc.
WCOM	MCI WorldCom Inc.
YHOO	Yahoo!, Inc.

not.

Additionally, we have to consider the determination of the inside quotes. Firstly, as the model is simulated, inside quotes do not enter exogenously anymore. They are now determined endogenously from the simulated quotes of the  $N$  included dealers (the highest bid of all dealers forms the inside bid and the lowest ask the inside ask). Second, since the quotes are issued at a discrete price grid we define a dealer to be at the inside if her quote is within the range of half a tick size from our simulated inside. This is done to compensate for the fact that in reality multiple dealers can be at the inside, whereas in our simulation, where prices take on any value, only one dealer can be at the inside.

When the baseline has been set conditional on the initial situation we can apply a one standard deviation shock with the size determined by our covariance matrix decomposition. The system is simulated again with the conditional bootstrap mentioned and the final IR function is calculated by subtracting the baseline from the shocked system.

### 3.3 Data

Nasdaq data is obtained from the Nasdaq data set provided by Nasdaq. This data set includes all transactions, dealer quotes and inside quotes issued at the Nasdaq trading system. Since this chapter focuses on dealer behavior we only consider the quote data in this chapter. One of the great advantages of this data set is that it provides time stamped quotes (to the



Table 3.2: Summary Statistics of the various companies and dealers:

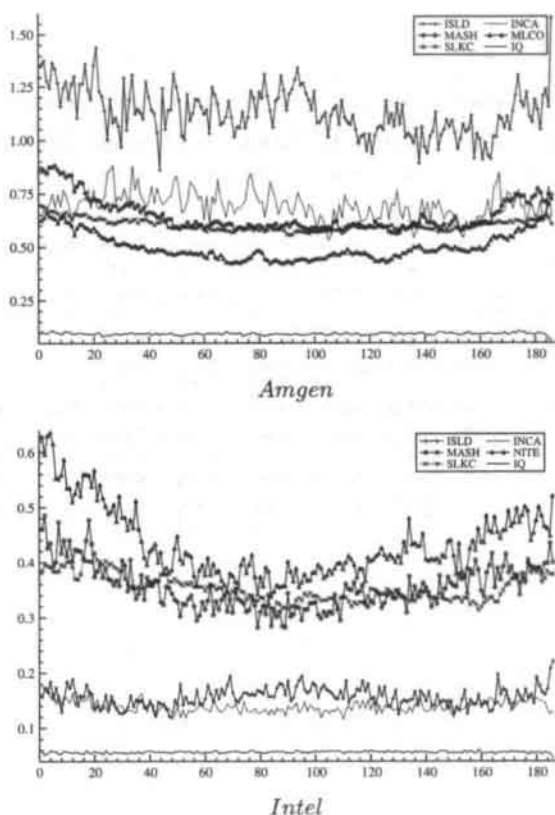
Stock	# of days	IQ	Island	Instinet	MM1	MM2	MM3
AAPL	124	0.0794	0.465	0.487	0.483	0.500	0.562
AMAT	124	0.0772	0.217	0.327	0.634	0.656	0.540
AMGN*	121	0.108	1.560	1.470	0.750	0.596	0.700
AMZN	120	0.145	0.357	0.832	1.296	4.411	1.742
ATHM*	122	0.176	0.729	1.836	1.180	2.316	1.904
CMGI*	123	0.257	0.654	3.485	2.534	1.838	2.564
COMS	123	0.0667	0.167	0.166	0.260	0.296	0.263
CPWR*	121	0.0831	0.623	1.613	0.570	0.848	0.992
CSCO*	123	0.0754	0.184	0.246	0.566	0.847	0.655
DELL*	123	0.0654	0.109	0.137	0.452	0.378	0.289
EGRP*	122	0.111	0.273	0.694	0.603	0.669	0.513
INTC*	123	0.0707	0.207	0.167	0.464	0.531	0.475
MSFT*	123	0.0722	0.187	0.212	0.480	0.535	0.449
NOVL	124	0.0687	0.306	0.627	0.305	0.413	0.256
NXTL	118	0.0760	0.898	1.345	0.505	0.637	0.379
ORCL	123	0.0666	0.142	0.246	0.336	0.439	0.235
PSFT	123	0.0683	0.209	0.284	0.276	0.359	0.270
SUNW*	123	0.0807	0.313	0.315	0.661	0.926	0.698
WCOM	124	0.0743	0.362	0.222	0.585	0.437	0.781
YHOO*	123	0.169	0.427	0.821	1.324	2.640	2.689

*Note:* This table reports average spreads for the firms our sample. Data used is from the period Januari 1999 to August 1999. The twenty stocks selected for our sample are named by their ticker symbol. Company names with an asterisk had a stock split occurring within our sample period. Number of days indicates the days without recording errors or other data problems, and thus indicates the days that where used for further analysis. For the inside quote and dealer quotes we report the average spreads (in absolute terms).

nearest second) together with the identity of the market maker.<sup>5</sup> Our sample runs from 1<sup>st</sup> February 1999 until 31<sup>st</sup> July 1999, giving us a total of 124 trading days. From the data set we select 20 companies that had the highest average trading volume over this period. Stock names and ticker symbols are reported in table 3.1. We use dealer quotes and inside quotes from twenty minutes after opening of the market (9.50 a.m.) until closing (4.00 p.m.). The first twenty minutes are left out in order to remove the strong diurnallity pattern that was observed during the opening of the market, and to circumvent the problem of a dealer not having posted a firm quote to the system yet. Dealers are ranked on their quoting frequency. From this ranking we extract the top two ECNs (Island and Instinet) and the top three market makers for each stock. For these five dealers (we refer to dealers being both ECNs and market makers) we observe quotes as well as inside quote (best bid and ask quote) at 120 second intervals. In this sense we deviate from Huang (2002), who pools dealers into groups. Our interest is in explaining individual dealer behavior. Moreover, Schultz (2003) shows that stocks are dominated by only few market makers, emphasizing the difference

<sup>5</sup>Huang (2002) uses the same data and provides an overview of the data set used.

Figure 3.1: Average spreads for Amgen and Intel per dealer



among individual market makers.

We remove all days where a dealer posts less than 30 quotes. Next we correct for stock splits by multiplying the data previous to the stock split with the ratio of the stock split, and include a dummy later on in our model for the stock split that has taken place. Finally, we compare all the dealer quotes with the inside quote and substitute all quotes that deviate more than \$ 5 from the inside quote.<sup>6</sup> We make this correction since a dealer always has to trade at the quotes she issued. Whenever a dealer does not want to trade, she will set her quote far away from the inside to prevent this quote from being hit. Since a non-willingness to trade does not add to the price discovery process, we do not want to take these quotes into our model. For zero quotes we follow the same adjustment procedure.

<sup>6</sup>These quotes are replaced by the last quote of the previous interval.

Table 3.3: Fraction of quotes at the inside for all dealers

Stock	Island		Instinet		MM1		MM2		MM3	
	bid	ask	bid	ask	bid	ask	bid	ask	bid	ask
AAPL	0.392	0.428	0.391	0.434	0.177	0.177	0.153	0.162	0.135	0.166
AMAT	0.594	0.658	0.408	0.417	0.105	0.111	0.082	0.079	0.184	0.158
AMGN	0.284	0.325	0.382	0.408	0.106	0.095	0.198	0.168	0.144	0.115
AMZN	0.543	0.597	0.287	0.276	0.048	0.045	0.029	0.035	0.085	0.059
ATHM	0.407	0.447	0.351	0.368	0.136	0.111	0.069	0.050	0.098	0.076
CMGI	0.516	0.550	0.212	0.216	0.099	0.094	0.142	0.136	0.071	0.061
COMS	0.572	0.598	0.608	0.610	0.330	0.269	0.288	0.240	0.346	0.261
CPWR	0.347	0.368	0.349	0.472	0.122	0.126	0.104	0.122	0.054	0.059
CSCO	0.607	0.646	0.519	0.494	0.143	0.105	0.131	0.117	0.068	0.048
DELL	0.708	0.744	0.637	0.646	0.277	0.239	0.275	0.281	0.221	0.182
EGRP	0.539	0.567	0.323	0.314	0.196	0.157	0.220	0.179	0.241	0.182
INTC	0.570	0.627	0.534	0.550	0.195	0.180	0.166	0.175	0.203	0.193
MSFT	0.595	0.640	0.574	0.574	0.142	0.134	0.132	0.112	0.139	0.170
NOVL	0.500	0.487	0.584	0.582	0.231	0.240	0.143	0.155	0.348	0.280
NXTL	0.284	0.346	0.534	0.570	0.137	0.170	0.123	0.155	0.213	0.199
ORCL	0.645	0.688	0.613	0.599	0.267	0.263	0.272	0.280	0.376	0.374
PSFT	0.498	0.465	0.491	0.507	0.368	0.286	0.303	0.209	0.321	0.302
SUNW	0.499	0.579	0.471	0.476	0.098	0.072	0.079	0.071	0.119	0.126
WCOM	0.468	0.541	0.542	0.540	0.125	0.109	0.213	0.176	0.065	0.084
YHOO	0.529	0.563	0.310	0.305	0.068	0.058	0.060	0.055	0.063	0.058

*Note:* This table reports the fraction of the total quotes that a dealer is at the inside.

In table 3.2 we report some summary statistics for our sample. From this table we see that spread patterns can differ quite a lot among the different stocks, where spreads are smaller for the market makers instead of the ECNs (AMGN) to being larger for market makers compared to ECNs (INTC).<sup>7</sup> This might indicate that although ECNs are taking over market share from Nasdaq, they still mainly focus on specific stocks traded at Nasdaq. Another interesting result is that, although the ECNs are in general most efficient with their quotes, there is still a big difference between their spreads and the inside spread. This is confirmed for the stocks of Amgen and Intel in figure 3.1, where we see that ECN spreads are relatively large compared to the spread for the inside quotes. The two stocks are selected because they represent two extremes in the sample. For Amgen market makers have the smaller spreads, for Intel ECNs have smaller spreads.

As we excluded the first 20 minutes of the day, we see that the intradaily pattern is quite stable and that we do not observe a strong U-shaped pattern over the day. This is in line with the findings of Chan, Christie, and Schultz (1995) who observe that the U-shaped pattern for Nasdaq stocks is much smaller than for NYSE listed stocks. From these graphs we can also infer that the inside quote is very stable over the day. Table 3.3 reports the fraction of times that dealers are at the inside. We find that overall ECNs are at the inside most of the

<sup>7</sup>The two stocks mentioned here will be used as examples throughout the chapter.

time, which is in line with Huang (2002).

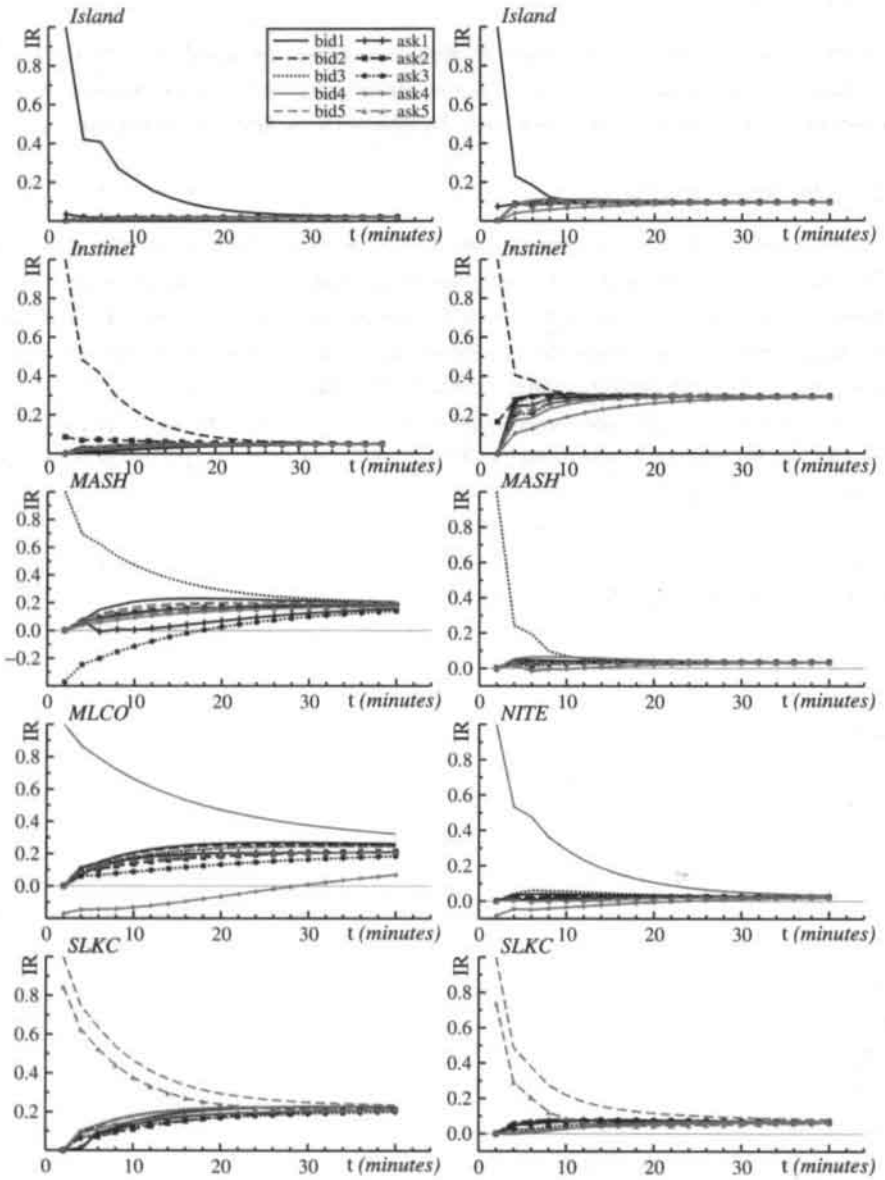
### 3.4 Results

In this section we present the results of the models and tests proposed in section 3.2. We first comment on the linear model as proposed by Hasbrouck (1991) and show linear impulse responses for the two stocks Aingen and Intel. Subsequently we address the nonlinear model.

#### 3.4.1 The linear model

As a benchmark for the nonlinear model, we first present the basic results for the linear VECM. Model (3.1) is estimated imposing the restrictions in (3.3) and excluding inside dummies. To investigate the quote dynamics we compute impulse responses. As mentioned linear impulse responses are insensitive to the shock size and history of the model. Hence, we apply a one standard deviation (unit) shock to the idiosyncratic part of each dealer in the system. However, we do take the correlations between bid and ask into account. These correlations are determined by the decomposition in (3.7), but we elaborate on the estimation of these at a later stage. Since shocks are standardized to a one standard deviation shock we cannot make any inferences on the absolute level of quote changes, but we do observe the relative change of the quotes due to a unit shock. Thus price discovery can be explained as the impact on quote changes due to a unit shock.

Figure 3.2: Linear Impulse response functions for Amgen and Intel: Shocks to bid



The impulse response functions are plotted in figure 3.2<sup>8</sup> for the two specific stocks Amgen (AMGN) and Intel (INTC). These graphs show how the impact of a unit shock to a dealer affects the quotes of the other dealers. A remarkable result is that these responses clearly relate to the ranking in spreads for dealers. For Amgen market makers have smaller spreads than the ECNs. In this case we observe that a shock to these dealers has a higher impact on the new level of the quotes than a shock to the ECNs. The opposite holds for Intel. For Amgen the three market makers move the quotes with about the same magnitude (MLCO has a slightly higher impact on the new level). For Intel it is obviously Instinet that has the highest impact on the other dealers' quotes. The implications from these graphs would indicate that price discovery and dealer efficiency would be one-on-one related to the size of the spreads that dealers quote.

The level of convergence is not the only relevant indicator of price discovery. Another relevant issue is the speed at which quotes converge to this level. This rate provides a good indication of how fast the price discovery and how quickly the information of a particular dealer is incorporated in the quotes of other dealers. If we consider the responses of the dealers to the shocks to Island and Instinet for Intel we see that all quotes have converged to about the same level within 20 minutes (recalling that we sample at a 2-minute interval). This indicates that private information is incorporated in the market within this time frame. For Amgen we see that after 40 minutes there are still differences in the levels that quotes correct to. Overall we can state that the information is incorporated into quotes more quickly for Intel than for Amgen. When we consider the three market makers for Amgen we observe that the private information of market maker 2 takes longer to be incorporated into quotes than for the other two market makers. Hence this dealer is considered less efficient than the others.

### 3.4.2 The Nonlinear Model

In this subsection we extend the linear model to a nonlinear model by including the inside quotes and the inside dummies. In table 3.4 we report the Wald statistics for the restricted versus the unrestricted model, as well as for the added coefficient matrices.

The first column of table 3.4 reports the test statistics for the restrictions in (3.3).<sup>9</sup> In 14 of the 20 cases we can reject the linear model at the 1% level and 16 at the 5% level. This clearly shows that the addition of inside quotes to the model leads to a significant improvement of the model in most of the cases. The other nonlinear term, the inside dummies are reported in the second column. These are all highly significant, with a  $\chi^2(100)$  at a 1% level of 136.

<sup>8</sup>These graphs do not show confidence intervals, since the point estimate and the confidence intervals deviate only marginally from each other and only blur the actual picture.

<sup>9</sup>These restrictions are tested with a  $\chi^2(20)$  for which the 5% and 1% confidence levels are at 31.4 and 37.6 respectively.

Table 3.4: Statistics for the Nonlinear model

Stock	Wald (lin-non)	$D_t^I$	$\Psi$	$\theta$	$\bar{R}^2$
AAPL	111.55	1238.0	3574.9	210.59	0.16
AMAT	57.30	1397.9	6009.8	256.66	0.18
AMGN	117.40	1690.2	4121.1	177.05	0.15
AMZN	34.85	480.83	7572.9	890.08	0.12
ATHM	37.61	605.32	7507.5	287.64	0.18
CMGI	67.58	580.54	16333	397.87	0.20
COMS	44.57	747.46	5142.1	268.01	0.17
CPWR	76.13	1133.2	2519.8	98.08	0.15
CSCO	22.36	980.68	9244.3	625.70	0.19
DELL	21.66	698.27	12800	1096.7	0.18
EGRP	39.56	343.79	9116.1	447.77	0.16
INTC	35.78	499.71	10362	365.13	0.18
MSFT	86.96	696.34	6826.9	863.29	0.18
NOVL	23.50	1174.2	2453.0	61.94	0.16
NXTL	72.23	1533.7	2706.3	70.57	0.15
ORCL	56.79	711.10	4694.2	258.11	0.13
PSFT	123.09	802.91	4054.0	114.79	0.18
SUNW	59.26	1378.7	7420.5	478.97	0.17
WCOM	29.97	1302.9	9958.9	285.20	0.21
YHOO	51.77	666.57	12822	928.64	0.14

*Note:* This table reports several statistics of the nonlinear model. The first column reports the Wald statistic of the restrictions that make the nonlinear model linear. The next three columns report the Wald statistics for the various coefficient matrices and vectors in the regression

$$\begin{bmatrix} \Delta b_t \\ \Delta a_t \end{bmatrix} = D_t(\cdot) + \Phi \begin{bmatrix} \Delta b_{t-1} \\ \Delta a_{t-1} \end{bmatrix} \\ + \Psi \begin{bmatrix} (b_{t-1} - i b i q_{t-1}) \\ (a_{t-1} - i a i q_{t-1}) \end{bmatrix} + \theta (a i q_{t-1} - b i q_{t-1}) + \begin{bmatrix} \eta_t \\ \nu_t \end{bmatrix}$$

The 1% level of a  $\chi^2$  with 20 degrees of freedom (for the restricted versus the unrestricted model) is at 37.6, for the Error correction term and the inside dummies with 100 degrees of freedom at 136 and for the inside spread with 10 degrees of freedom at 23.2. The last column reports the average R-squared of the estimated model.

Table 3.5: Matrix of Median coefficients for Inside Dummies

	Island		Instinet		MM1		MM2		MM3	
ISLD $\Delta b_t$	0	-0.026	0.021	-0.027	-0.002	0	0.001	0.006	-0.006	-0.001
	(12)	(13)	(11)	(13)	(1)	(1)	(1)	(0)	(2)	(1)
ISLD $\Delta a_t$	0.023	-0.013	0.014	-0.019	0.003	0.003	0.002	0.001	-0.002	0.002
	(12)	(11)	(11)	(8)	(0)	(1)	(0)	(0)	(0)	(1)
INST $\Delta b_t$	0.031	-0.034	-0.038	-0.024	-0.002	-0.006	0.004	0.002	-0.003	0.001
	(12)	(15)	(13)	(10)	(0)	(0)	(1)	(0)	(1)	(1)
INST $\Delta a_t$	0.028	-0.03	0.014	0.021	0.003	-0.003	0.001	0	-0.007	0.005
	(15)	(11)	(8)	(10)	(0)	(1)	(0)	(0)	(0)	(0)
MM1 $\Delta b_t$	0.026	-0.028	0.016	-0.012	-0.051	-0.013	0.002	0.002	-0.007	0.006
	(15)	(16)	(10)	(6)	(14)	(4)	(3)	(0)	(2)	(1)
MM1 $\Delta a_t$	0.026	-0.017	0.014	-0.012	0.009	0.047	0.003	0.001	-0.002	0.01
	(14)	(10)	(6)	(5)	(4)	(16)	(0)	(1)	(0)	(1)
MM2 $\Delta b_t$	0.022	-0.027	0.016	-0.015	0.005	-0.001	-0.075	0.003	-0.003	0.006
	(12)	(15)	(9)	(8)	(0)	(0)	(18)	(3)	(1)	(0)
MM2 $\Delta a_t$	0.021	-0.022	0.011	-0.018	0.003	-0.005	0.006	0.085	-0.006	0.005
	(12)	(13)	(5)	(9)	(1)	(0)	(4)	(19)	(0)	(1)
MM3 $\Delta b_t$	0.034	-0.037	0.02	-0.019	-0.001	0.001	0.002	0.001	-0.063	0.022
	(18)	(19)	(13)	(14)	(0)	(1)	(5)	(0)	(18)	(12)
MM3 $\Delta a_t$	0.033	-0.036	0.019	-0.017	-0.004	0.003	0.001	0	-0.03	0.08
	(18)	(19)	(14)	(13)	(0)	(0)	(1)	(0)	(13)	(20)

Note: This Table reports the coefficients on the Inside Dummies in the nonlinear model:

$$\begin{bmatrix} \Delta b_t \\ \Delta a_t \end{bmatrix} = D_t(\cdot) + \Phi \begin{bmatrix} \Delta b_{t-1} \\ \Delta a_{t-1} \end{bmatrix} + \Psi \begin{bmatrix} (b_{t-1} - i b i q_{t-1}) \\ (a_{t-1} - i a i q_{t-1}) \end{bmatrix} \\ + \theta (a i q_{t-1} - b i q_{t-1}) + \begin{bmatrix} \eta_t \\ \nu_t \end{bmatrix}$$

The inside dummies are included in the deterministic function  $D_t^I$ . The medians of all 20 stocks in the sample are reported. In parentheses we report the number of firms for which the coefficient is significant at the 1% level in the model.

The next two columns report the Wald statistic for the two coefficient matrices  $\Psi$  and vector  $\theta$ , which are all significant as well. The results in this table provides us with clear evidence that the nonlinearities are highly significant with the greatest improvement of the model lies in the addition of the inside dummies. The last column of this table reports the average R-squared per stock. On average the fit of the model ranges between 0.11 and 0.21. This fit also differs per dealers (not reported), where the average fit for Island equals 0.18, for Instinet 0.19, MM1 0.22 and MM2 and MM3 both 0.13.

Table 3.5 presents the coefficients on the lagged inside dummies. Focussing on the diagonal of this matrix we see an increasing number of significant coefficients when we go down the diagonal. The sign of these coefficients is undetermined for the bid of Island, but negative for the ask. This means that if Island has the inside ask it will lower the ask quote even further. For the other dealers the coefficients for bids are negative and positive for the asks, indicating a movement away from the inside, when at the inside. There are a couple of other interesting results observed in this table. First, there is a positive reaction of other dealers' quotes to



Table 3.6: Matrix of Median coefficients for lagged quotes  $\Phi$ 

	Island		Instinet		MM1		MM2		MM3	
	$\Delta b_{t-1}$	$\Delta a_{t-1}$	$\Delta b_{t-1}$	$\Delta a_{t-1}$	$\Delta b_{t-1}$	$\Delta a_{t-1}$	$\Delta b_{t-1}$	$\Delta a_{t-1}$	$\Delta b_{t-1}$	$\Delta a_{t-1}$
ISLD $\Delta b_t$	-0.101 (18)	0.038 (15)	0.004 (0)	0.013 (2)	-0.001 (0)	0.001 (0)	-0.003 (1)	0.01 (0)	-0.001 (0)	0.015 (0)
ISLD $\Delta a_t$	0.033 (13)	-0.098 (15)	0.013 (4)	-0.004 (1)	0.008 (1)	-0.011 (2)	0.004 (0)	-0.001 (0)	0.004 (0)	-0.003 (0)
INST $\Delta b_t$	0.008 (1)	0.027 (10)	-0.129 (17)	0.015 (3)	0.006 (0)	0.01 (2)	0.002 (0)	0.009 (1)	0.001 (0)	0.016 (1)
INST $\Delta a_t$	0.022 (11)	0.007 (3)	0.016 (2)	-0.133 (19)	0.015 (6)	-0.011 (0)	0.011 (1)	-0.004 (0)	0.014 (0)	0.002 (1)
MM1 $\Delta b_t$	0 (1)	0.015 (5)	0.006 (1)	0.014 (6)	-0.105 (20)	0.024 (6)	0 (5)	0.015 (2)	-0.004 (2)	0.02 (2)
MM1 $\Delta a_t$	0.014 (5)	-0.002 (1)	0.012 (5)	0.003 (1)	0.025 (8)	-0.127 (20)	0.011 (3)	-0.008 (2)	0.029 (7)	0.002 (2)
MM2 $\Delta b_t$	0.01 (2)	0.032 (8)	0.012 (2)	0.017 (7)	0.008 (2)	0.005 (1)	-0.108 (19)	0.024 (7)	0.013 (3)	0.003 (0)
MM2 $\Delta a_t$	0.02 (9)	0.01 (4)	0.018 (8)	0.003 (2)	0.015 (5)	-0.005 (1)	0.018 (5)	-0.116 (19)	0.021 (3)	-0.004 (1)
MM3 $\Delta b_t$	0.016 (5)	0.028 (10)	0.014 (8)	0.016 (7)	0.019 (6)	0.011 (2)	0.005 (4)	0.006 (1)	-0.097 (15)	0.019 (3)
MM3 $\Delta a_t$	0.018 (8)	0.023 (6)	0.015 (6)	0.011 (4)	0.019 (7)	-0.002 (2)	0.007 (4)	0.005 (2)	0.027 (6)	-0.096 (17)

Note: This table reports the coefficients of the  $\Phi$  matrix in the nonlinear model:

$$\begin{bmatrix} \Delta b_t \\ \Delta a_t \end{bmatrix} = D_t(\cdot) + \Phi \begin{bmatrix} \Delta b_{t-1} \\ \Delta a_{t-1} \end{bmatrix} + \Psi \begin{bmatrix} (b_{t-1} - i \text{ bi}q_{t-1}) \\ (a_{t-1} - i \text{ ai}q_{t-1}) \end{bmatrix} \\ + \theta(\text{ai}q_{t-1} - \text{bi}q_{t-1}) + \begin{bmatrix} \eta_t \\ \nu_t \end{bmatrix}.$$

In the top row  $\Delta b_{t-1}$  refers to  $\Delta \text{bid}_{t-1}$  and  $\Delta a_{t-1}$  to  $\Delta \text{ask}_{t-1}$ . The medians of all 20 stocks in the sample are reported. In parentheses we report the number of firms for which  $\phi_{ij}$  is significant at the 1% level in the model.

Island being at the inside bid and how their adjustment is done symmetrically. Dealers tend to raise both quotes with the same amount when Island is at the inside bid. A similar result is obtained when Island is at the inside ask, all other dealers tend to lower their quotes, again in a symmetric fashion. This is an indication of the importance of this ECN. For Instinet a similar conclusion can be drawn although these results are less significant. On the other side of the spectrum we observe the reaction of the ECNs to market makers being at the inside. There is only a very small reaction of ECNs to market makers being at the inside, with only few significant coefficients.

In table 3.6 we present the coefficients of the AR(1) component. We observe that all coefficients along the diagonal are negative. Hence, when a dealer e.g. raises a quote she is more inclined to lower that quote again in the next period. We find the magnitude of this effect to be the same for all dealers. Furthermore there are some cross effects present between the bid and ask of the same dealer, most of the other cross effects are insignificant.

Table 3.7 presents the coefficients of the equilibrium correction term. Again we observe

Table 3.7: Matrix of Median coefficients for Error Correction term  $\Psi$ 

	Island		Instinet		MM1		MM2		MM3	
ISLD $\Delta b_t$	-0.604 (20)	-0.01 (5)	0.021 (9)	0.013 (1)	0.031 (11)	0.01 (3)	0.014 (5)	0.003 (0)	0.011 (4)	-0.005 (2)
ISLD $\Delta a_t$	-0.006 (5)	-0.63 (20)	0.009 (2)	0.034 (9)	0.006 (1)	0.047 (17)	0.001 (3)	0.024 (10)	-0.002 (3)	0.018 (4)
INST $\Delta b_t$	0.024 (3)	0.006 (1)	-0.53 (20)	0.002 (4)	0.019 (2)	0 (2)	0.012 (6)	-0.004 (0)	0.018 (5)	-0.016 (2)
INST $\Delta a_t$	-0.004 (1)	0.036 (8)	0.007 (2)	-0.486 (20)	-0.006 (0)	0.045 (14)	-0.001 (1)	0.017 (9)	-0.013 (3)	0.021 (8)
MM1 $\Delta b_t$	0.033 (12)	0.009 (3)	0.028 (7)	0.003 (3)	-0.553 (20)	-0.021 (6)	0.023 (13)	-0.004 (2)	0.017 (9)	-0.011 (3)
MM1 $\Delta a_t$	0.002 (2)	0.069 (19)	0.002 (0)	0.04 (15)	-0.03 (8)	-0.435 (20)	0 (6)	0.03 (16)	-0.024 (11)	0.03 (11)
MM2 $\Delta b_t$	0.018 (5)	0.003 (0)	0.011 (4)	0.004 (1)	0.012 (2)	-0.004 (4)	-0.241 (19)	-0.021 (9)	0.01 (3)	-0.006 (3)
MM2 $\Delta a_t$	0.005 (1)	0.025 (8)	0 (0)	0.015 (6)	-0.012 (4)	0.024 (9)	-0.023 (11)	-0.208 (20)	-0.008 (5)	0.013 (7)
MM3 $\Delta b_t$	0.021 (3)	0.009 (2)	0.01 (3)	0.005 (1)	0 (5)	-0.009 (5)	-0.002 (4)	-0.002 (0)	-0.215 (20)	-0.086 (16)
MM3 $\Delta a_t$	0.003 (1)	0.012 (3)	0.002 (1)	0.016 (6)	-0.009 (1)	0.009 (7)	-0.003 (3)	0.007 (6)	-0.079 (18)	-0.199 (20)

Note: This table reports the coefficients of the  $\Psi$  matrix in the nonlinear model:

$$\begin{bmatrix} \Delta b_t \\ \Delta a_t \end{bmatrix} = D_t(\cdot) + \Phi \begin{bmatrix} \Delta b_{t-1} \\ \Delta a_{t-1} \end{bmatrix} + \Psi \begin{bmatrix} b_{t-1} - \iota biq_{t-1} \\ a_{t-1} - \iota aiq_{t-1} \end{bmatrix} \\ + \theta(aiq_{t-1} - biq_{t-1}) + \begin{bmatrix} \eta_t \\ \nu_t \end{bmatrix}$$

The medians of all 20 stocks in the sample are reported. In parentheses we report the number of firms for which  $\psi_{i,j}$  is significant at the 1% level in the model.

a strong effect along the diagonal of the matrix, where all coefficients are negative and significant. Hence all dealers have a strong correction towards the respective inside quotes. Note that these coefficients get smaller in absolute terms as we go down the diagonal. This is in line with the fractions of time a dealer is at the inside as we observed in table 3.3 and indicates that ECNs have a stronger error correction than market makers. There seems to be no cross effects between bid and ask quotes of the same dealer, except for market maker 3, who has her other quote moving away from the inside.<sup>10</sup> This indicates that dealers only focus on the respective inside quote and do not consider the other inside quote. All other cross effects are mostly insignificant.

Table 3.8 reports the coefficients on the inside spread. There is a clear adjustment present to keep the spreads small as can be seen from the signs of the coefficients. This adjustment is stronger for the ECNs and less pronounced for the market makers. This finding is in line with the fact that ECNs profit from maintaining small spreads, whereas market makers could potentially profit from larger spreads. Hence ECNs are more inclined to keep the spreads

<sup>10</sup>Market maker 3 is often represented by the same market maker SLKC.

Table 3.8: Median Coefficients for the Inside Spread

	$aiq_{t-1} - biq_{t-1}$	
ISLD $\Delta b_t$	0.305	17
ISLD $\Delta a_t$	-0.135	10
INST $\Delta b_t$	0.340	19
INST $\Delta a_t$	-0.164	12
MM1 $\Delta b_t$	0.205	17
MM1 $\Delta a_t$	-0.066	5
MM2 $\Delta b_t$	0.170	13
MM2 $\Delta a_t$	-0.026	2
MM3 $\Delta b_t$	0.121	7
MM3 $\Delta a_t$	-0.0039	5

*Note:* This table reports the coefficients of the  $\theta$  vector in the nonlinear model:

$$\begin{bmatrix} \Delta b_t \\ \Delta a_t \end{bmatrix} = D_t(\cdot) + \Phi \begin{bmatrix} \Delta b_{t-1} \\ \Delta a_{t-1} \end{bmatrix} + \Psi \begin{bmatrix} (b_{t-1} - \iota biq_{t-1}) \\ (a_{t-1} - \iota aiq_{t-1}) \end{bmatrix} \\ + \theta(aiq_{t-1} - biq_{t-1}) + \begin{bmatrix} \eta_t \\ \nu_t \end{bmatrix}$$

In the first column the medians of all 20 stocks in the sample are reported. In the second column we report the number of firms for which this term is significant at the 1% level.

small than market makers (of course competition between market makers, induces market makers to achieve small spreads as well). Another interesting result is that the coefficients on the changes in bids are always larger, in absolute terms, than the coefficients of the changes in the ask. Hence there is an asymmetric adjustment of dealer quotes to the size of the spread. Dealers are more inclined to raise the bid quote than to lower the ask quote.

The results presented above all indicate that dealer bid and ask quotes display nonlinear features, with the strongest nonlinearity being the fact whether a dealer is at the inside or not. As this nonlinearity influences the dynamics of dealer quotes we are interested in whether it also affects second order moments. We perform a Breusch-Pagan test on (3.5) and (3.6) to test for regime dependent heteroskedasticity. In table 3.9 we present the results of the Breusch-Pagan test. This is a  $\chi^2$  test with the degrees of freedom equal to the number of regressors minus 1, excluding the constant. We report the number of regressions for which the test was significant at the 1 % level. The diagonal of this matrix represent the variance of each dealer's bid or ask quote and the off-diagonal the covariances. The results of this test remain rather inconclusive. In about 50% of the cases we find a significant relationship between the variance of quote changes and being at the inside. For the off-diagonal elements we find even less significance. We remain inconclusive in this, but note that for some stocks in the sample this effect has a significant impact on the behavior of dealers.

As the results of this regime dependent heteroskedasticity remain inconclusive, we can motivate the decomposition proposed in (3.7). If regime dependent heteroskedasticity was present, the decomposition in (3.7) has to take this into account and decompose the covariance

Table 3.9: Breusch Pagan test for Regime dependent Heteroscedasticity:

-	$\eta^1$	$\nu^1$	$\eta^2$	$\nu^2$	$\eta^3$	$\nu^3$	$\eta^4$	$\nu^4$	$\eta^5$	$\nu^5$
$\eta^1$	12	-	-	-	-	-	-	-	-	-
$\nu^1$	6	12	-	-	-	-	-	-	-	-
$\eta^2$	8	9	6	-	-	-	-	-	-	-
$\nu^2$	5	7	2	9	-	-	-	-	-	-
$\eta^3$	7	7	4	3	11	-	-	-	-	-
$\nu^3$	10	9	4	5	4	8	-	-	-	-
$\eta^4$	10	8	5	4	6	5	11	-	-	-
$\nu^4$	8	7	6	7	5	5	6	10	-	-
$\eta^5$	11	11	8	7	9	9	8	7	11	-
$\nu^5$	11	10	8	7	8	10	7	7	9	10

*Note:* This table reports the number of regressions for which we found regime dependent heteroskedasticity at the 1% significance level. The diagonal of the table show this number for the variances using the regression

$$\Omega_{ii,t} = \gamma_{0,i} + \gamma_{1,i} D_{b,i,t}^I + \gamma_{2,i} D_{a,i,t}^I,$$

where  $\Omega_{ii,t}$  is the bid or ask variance of dealer  $i$ ,  $D_{b,i,t}^I$  is the inside dummy equal to 1 if the bid of dealer  $i$  is at the inside and  $D_{a,i,t}^I$  the inside dummy equal to 1 if the ask of dealer  $i$  is at the inside. The elements in the lower off-diagonal show this number for the covariances, using the regression

$$\Omega_{ij,t} = \lambda_{0,i,j} + \lambda_{1,i,j} D_{b,i,t}^I + \lambda_{2,i,j} D_{a,i,t}^I + \lambda_{3,i,j} D_{b,j,t}^I + \lambda_{4,i,j} D_{a,j,t}^I,$$

where  $\Omega_{ij,t}$  is the covariance between a particular quote of dealer  $i$  and dealer  $j$ .

matrix conditional on the state that it is in. We impose the structure of the covariance matrix as in (3.8) on the empirical covariance matrix that we obtain from the regression. We use GMM to achieve this, by fitting the parameters in our restricted covariance matrix to the unrestricted covariance matrix.<sup>11</sup> Panel A in tables 3.10 and 3.11 show the structure of the covariance matrix obtained from the regression. Panel B shows the restricted structure of this covariance matrix.

When considering table 3.10 we observe that the noise terms for the ECNs are much larger than for the market makers. In table 3.3 we already observed that these ECNs had the larger spreads. Clearly there is a lot of idiosyncratic noise around the dynamics for these ECNs. Another interesting result is the cross-dynamics observed between ECNs. Although small compared to the variances, they are still twice the size of the fundamental part. A final remark to this table is the correlation between the bid and ask quote for market maker 3. These error components are almost perfectly correlated. Turning to the covariance matrix

<sup>11</sup>i.e. setting the orthogonality conditions  $E[\tilde{\Omega} - \Omega] = 0$ , where  $\tilde{\Omega}$  is the restricted covariance matrix as defined in 3.8 and  $\Omega$  is the unconditional covariance matrix.

Table 3.10: Covariance Matrix of the Nonlinear model for Amgen

Panel A: Unrestricted Covariance Matrix									
1.224	0.072	0.073	0.073	0.036	0.036	0.037	0.036	0.053	0.052
0.072	0.864	0.062	0.060	0.036	0.032	0.035	0.031	0.043	0.043
0.073	0.062	0.587	0.082	0.036	0.029	0.034	0.033	0.046	0.044
0.073	0.060	0.082	0.473	0.032	0.030	0.032	0.032	0.043	0.044
0.036	0.036	0.036	0.032	0.070	0.026	0.031	0.027	0.034	0.033
0.036	0.032	0.029	0.030	0.026	0.073	0.026	0.028	0.030	0.030
0.037	0.035	0.034	0.032	0.031	0.026	0.053	0.035	0.037	0.036
0.036	0.031	0.033	0.032	0.027	0.028	0.035	0.058	0.034	0.034
0.053	0.043	0.046	0.043	0.034	0.030	0.037	0.034	0.088	0.080
0.052	0.043	0.044	0.044	0.033	0.030	0.036	0.034	0.080	0.087
Panel B: Restricted Covariance Matrix									
1.224	0.072	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039
0.072	0.864	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039
0.039	0.039	0.587	0.082	0.039	0.039	0.039	0.039	0.039	0.039
0.039	0.039	0.082	0.473	0.039	0.039	0.039	0.039	0.039	0.039
0.039	0.039	0.039	0.039	0.070	0.026	0.039	0.039	0.039	0.039
0.039	0.039	0.039	0.039	0.026	0.073	0.039	0.039	0.039	0.039
0.039	0.039	0.039	0.039	0.039	0.039	0.053	0.035	0.039	0.039
0.039	0.039	0.039	0.039	0.039	0.039	0.035	0.058	0.039	0.039
0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.088	0.080
0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.080	0.087

Note: This table provides the unrestricted and restricted covariance matrix for the stock Amgen. The unrestricted covariance matrix is decomposed into a common factor (market wide news) and idiosyncratic dealer components. The decomposition has the following form:

$$\eta_t = \varepsilon_t \iota + \xi_{b,t}, \quad \nu_t = \varepsilon_t \iota + \xi_{a,t}$$

where  $\eta_t$  and  $\nu_t$  are the innovation terms of the nonlinear model,  $\varepsilon_t$  represents the market noise and  $\xi_{b,t}$  and  $\xi_{a,t}$  represent the dealers' idiosyncratic noise.

Table 3.11: Covariance Matrix of the Nonlinear model for Intel

Panel A: Unrestricted Covariance Matrix									
0.082	0.039	0.041	0.039	0.043	0.037	0.035	0.032	0.04	0.039
0.039	0.06	0.037	0.038	0.039	0.036	0.031	0.032	0.038	0.038
0.041	0.037	0.051	0.038	0.041	0.035	0.034	0.032	0.04	0.039
0.039	0.038	0.038	0.05	0.04	0.036	0.032	0.033	0.039	0.038
0.043	0.039	0.041	0.04	0.101	0.036	0.038	0.033	0.042	0.04
0.037	0.036	0.035	0.036	0.036	0.089	0.029	0.034	0.036	0.036
0.035	0.031	0.034	0.032	0.038	0.029	0.123	0.029	0.034	0.033
0.032	0.032	0.032	0.033	0.033	0.034	0.029	0.129	0.034	0.034
0.04	0.038	0.04	0.039	0.042	0.036	0.034	0.034	0.076	0.065
0.039	0.038	0.039	0.038	0.04	0.036	0.033	0.034	0.065	0.074
Panel B: Restricted Covariance Matrix									
0.082	0.039	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
0.039	0.06	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
0.036	0.036	0.051	0.038	0.036	0.036	0.036	0.036	0.036	0.036
0.036	0.036	0.038	0.05	0.036	0.036	0.036	0.036	0.036	0.036
0.036	0.036	0.036	0.036	0.101	0.036	0.036	0.036	0.036	0.036
0.036	0.036	0.036	0.036	0.036	0.089	0.036	0.036	0.036	0.036
0.036	0.036	0.036	0.036	0.036	0.036	0.123	0.029	0.036	0.036
0.036	0.036	0.036	0.036	0.036	0.036	0.029	0.129	0.036	0.036
0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.076	0.065
0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.065	0.074

*Note:* This table provides the unrestricted and restricted covariance matrix for the stock Intel. The unrestricted covariance matrix is decomposed into a common factor (market wide news) and idiosyncratic dealer components. The decomposition has the following form:

$$\eta_t = \varepsilon_t \iota + \xi_{b,t}, \quad \nu_t = \varepsilon_t \iota + \xi_{a,t}$$

where  $\eta_t$  and  $\nu_t$  are the innovation terms of the nonlinear model,  $\varepsilon_t$  represents the market noise and  $\xi_{b,t}$  and  $\xi_{a,t}$  represent the dealers' idiosyncratic noise.

of Intel in table 3.11 we observe the same thing for market maker 3. However, for this stock the market makers tend to have a higher variance than the ECNs. Further, the cross-effects between ECNs have disappeared.

The fundamental noise in the market concerns the information common to all dealers. This can be considered as market wide information (public information) but is also due to time aggregation of dealer quotes (see Hasbrouck (1995)). With the exception of market maker 3 all correlations between private information in bid and ask quotes of individual dealers are close to zero. This is in line with Jang and Venkatesh (1991), who finds that in most cases when dealers receive private information, they will only adjust their quote on one side of the market.

Although being very informative on the nature of the dealer, the main use of this decomposition is to provide us with parameters that indicate the size of the fundamental news and the idiosyncratic part, individual to every dealer. We need the sizes of these individual parts in order to apply shocks to the nonlinear model.

### 3.4.3 Non-Linear Impulse Responses

After having determined reasonable shock sizes, we can now conduct experiments on dealer quote reactions to dealer specific shocks. As mentioned in section 3.2 we use Generalized Impulse Response Functions, as the model is nonlinear. We simulate the model 5000 times and use a time horizon of 20 steps which, in our model, represents 40 minutes. We use a conditional bootstrap, where we draw from the innovation term in (3.1) in such a way that we maintain positive spreads. The inside quotes in this simulation are determined as the highest bid of the dealers evaluated and the lowest ask of the dealers. As quotes can take on any value in this simulation, there can only be one dealer having the inside. To allow for multiple dealers at the inside, we let all dealers within half a tick size from the best quote be at the inside as well. This entails that the inside dummy is equal to one for all dealers at the inside or within half a tick size away from the inside. Note that in this simulation there will always be a dealer at the inside, which need not be the case when the inside quotes are determined exogenously.

The first step in the process is to determine the history of the model. This entails that we set an initial situation in which the model starts (e.g. all dealers are at the inside, or at average spreads). Given this initial situation we simulate the model and find an average baseline for how the model will react on this initial situation. In the second step we apply a shock with a predetermined size to a particular dealer. Conditional on the same history the model is simulated again. Similar to the simulation of the baseline, we compute the average price path. The difference between the simulation including the shock model and the baseline yields the impulse response function.

We consider four experiments of interest. The first experiment we are interested in, is a one standard deviation shock in dealer's idiosyncratic noise (information), towards the inside quote, when all dealers start at average spreads. The results of these impulse response functions are plotted in figure 3.3. In this figure shocks are standardized and are thus comparable to the linear impulse responses. We only show shocks for the bid quotes as the plots for the asks show similar patterns. When comparing the nonlinear impulse response function to the linear impulse responses, there are a few differences observed. Firstly, for Amgen we see that the response of dealer quotes are larger than in the linear case, i.e. dealers react more to the shocks of other dealers. In the linear case we observed that all market makers had about the same impact on the quote level. In the nonlinear case we see that the second market maker causes the greatest change in this level. Among the three market makers, the lowest impact is caused by the last market maker. Furthermore, Instinet ECN has about the same impact on the quote level as the last market maker. For Intel we also observe some interesting features. In the linear case we saw that Instinet had the highest impact on the quote level. In the nonlinear case this is Island. The impact the market makers have on the quote level remains marginal, where we actually observe a slight decrease in the quote level.

We consider the next two experiments jointly as they do not lead to any significant differences from the previous situation. The first experiment we consider is a one standard deviation shock inwards, where the particular dealer starts at the inside, all other dealers are at average spreads. The second experiment has the same initial situation as the first experiment but now we apply a shock outwards. Figures 3.4 and 3.5 show the plots of these functions. There are some differences between the outcomes of the functions compared to the previous nonlinear impulse response function, but these are all marginal. Hence we conclude that the initial situation does not influence the outcome of the function that much, and improving the inside or leaving the inside quote does not reveal any additional information.

The last experiment that we consider is a situation where all dealer start at average spread, where we apply a shock such that the respective dealer reaches the inside. In this case we use the correlations that we observed between the private information component to determine the appropriate shock size for the other quote of the dealer. The initial situation is equal to the situation for the first experiment, but the shock size differs. The plots of the impulse response functions are shown in figure 3.6. In this case we do observe some clear differences with the impulse responses in figure 3.3. Firstly the impact that the market makers have on the quote level for Amgen is lower in this experiment. Moreover the impact of Instinet even increases further and has a higher impact on the quote level than the last market maker. The last market maker has an impact approximately equal to the impact Island has. For Intel we only observe some marginal differences.



Figure 3.3: Nonlinear Impulse response functions for Amgen (left) and Intel (right), starting at average spreads, shock inwards.

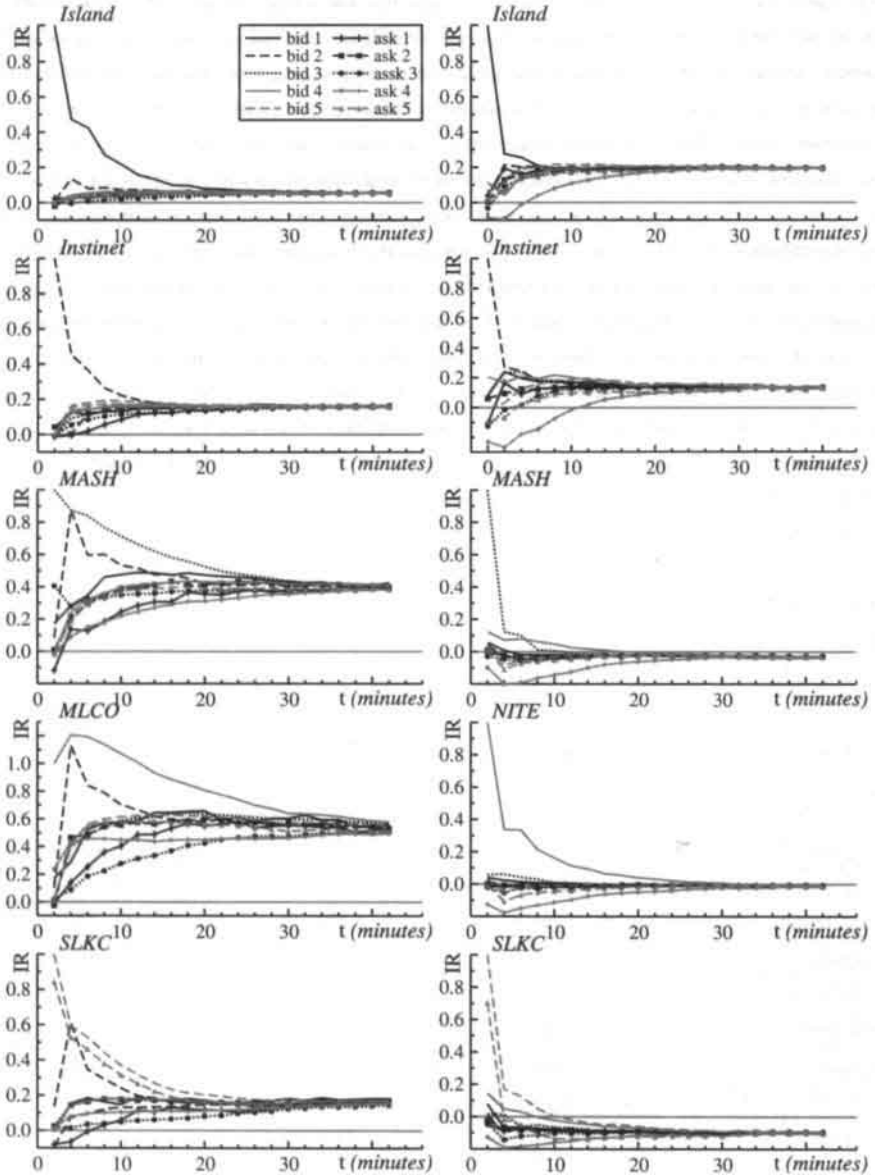


Figure 3.4: Nonlinear Impulse response functions for Amgen (left) and Intel (right), starting at inside for dealer, shock inwards.

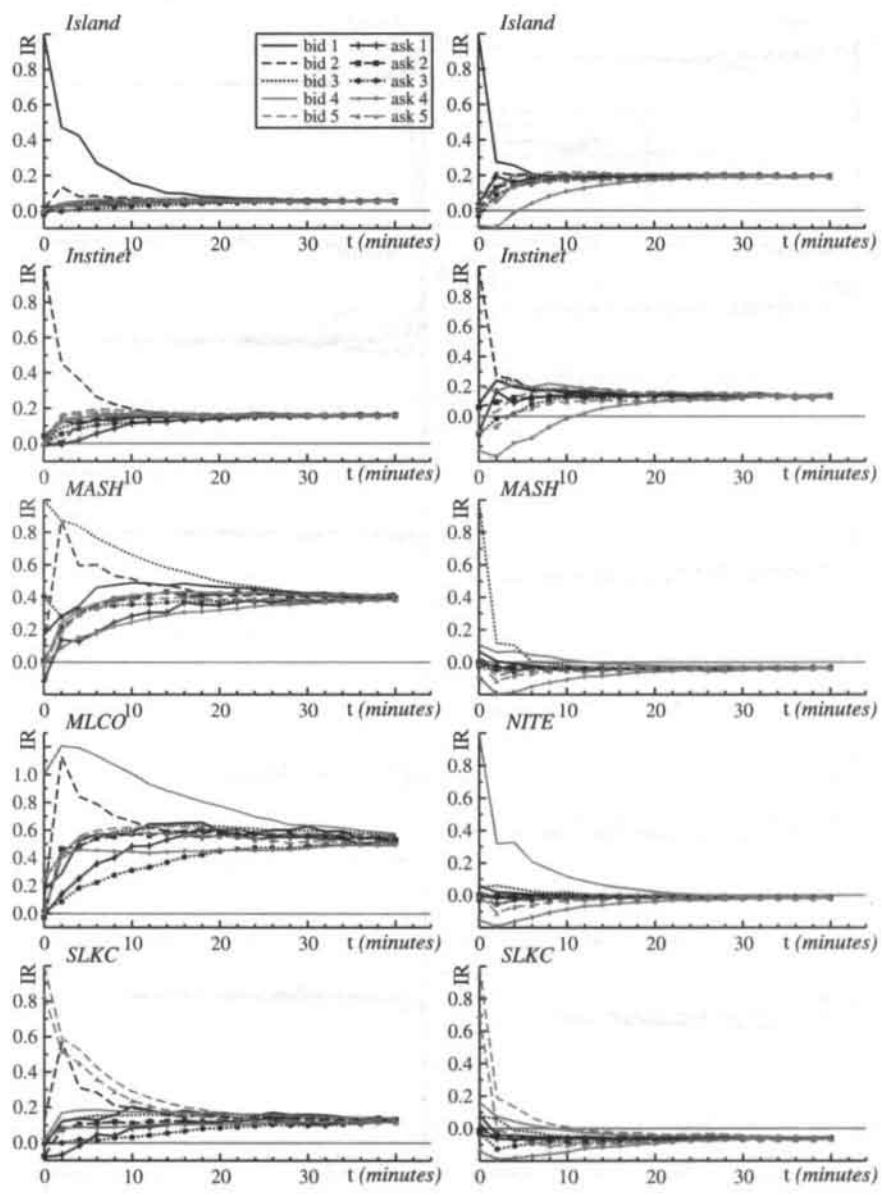


Figure 3.5: Nonlinear Impulse response functions for Amgen (left) and Intel (right), starting at inside for dealer, shock outwards.

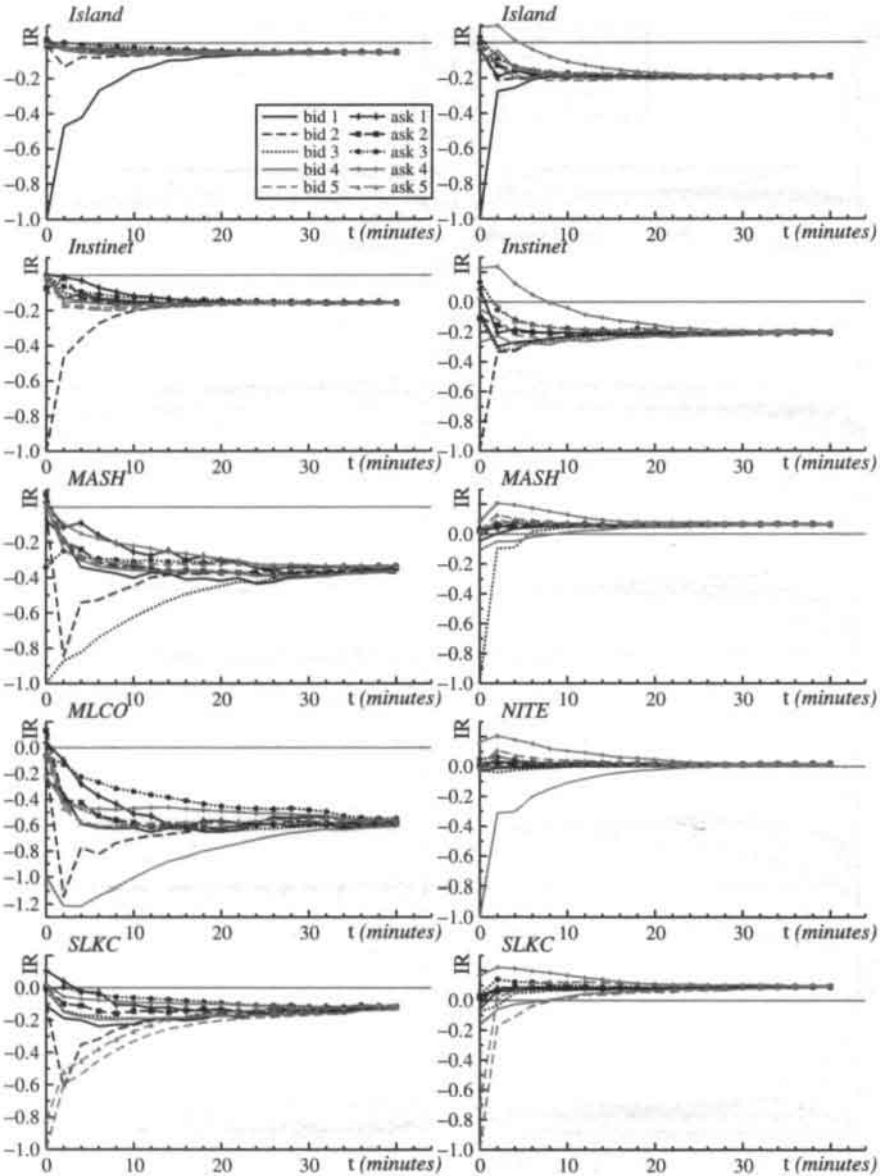
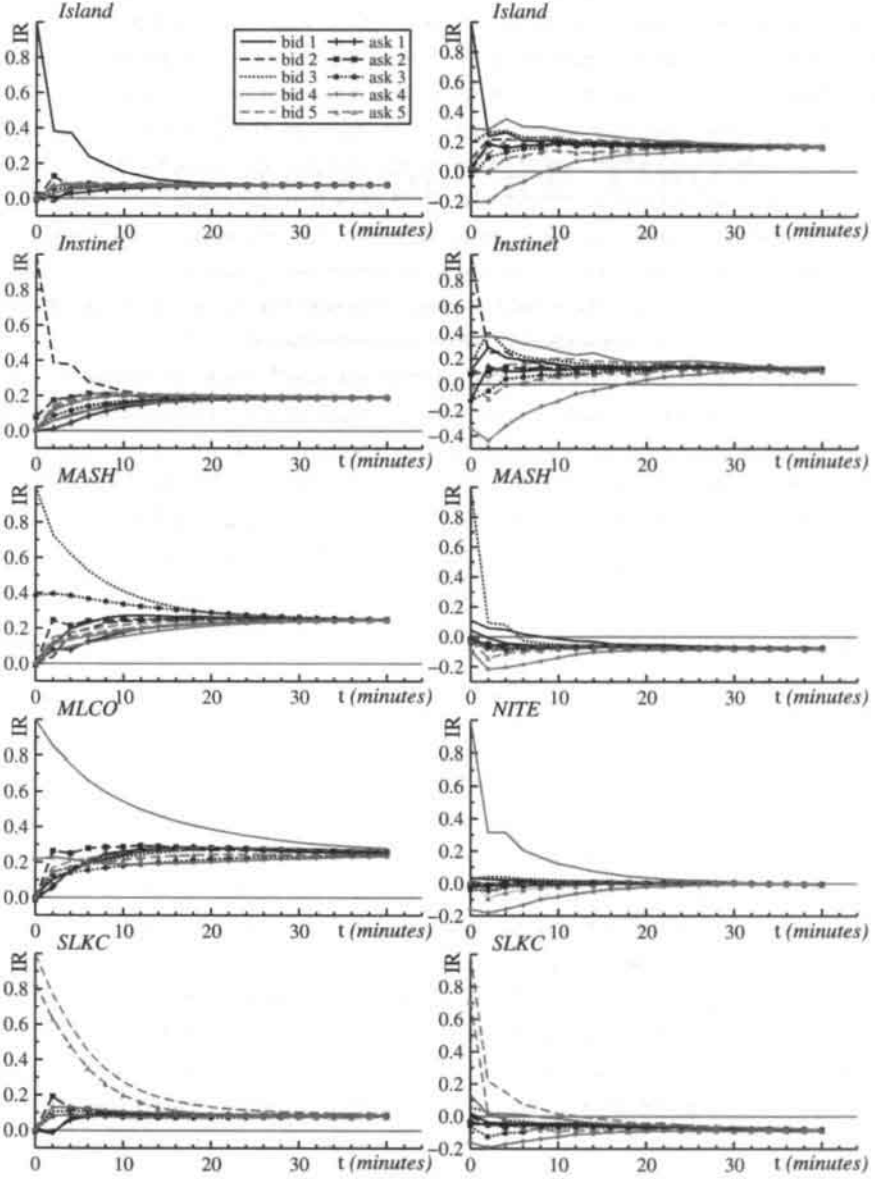


Figure 3.6: Nonlinear Impulse response functions for Amgen (left) and Intel (right), starting at mean spread, shock reaching inside



### 3.5 Conclusion

In this chapter we propose a nonlinear error correction model for quote changes. This model is an extension of the the linear model of Hasbrouck (1995). The nonlinearities are that we allow dealer quotes to correct towards the inside quotes and we also allow them to react to the situation dealers find themselves in (i.e. being at the inside). These nonlinearities are also included in second moments to test for regime dependent heteroskedasticity.

Overall, we find that our model significantly improves the linear model. Specifically, our results include that all dealers strongly correct towards the inside quotes. However, all dealers, except Island, have the tendency to move away from the inside, when they are at the inside. We further observe that all dealer move towards the inside quote when Island ECN is at the inside. Dealers react asymmetrically to the size of the spread, where they tend to raise their bid quote more, when the spread is large, than that they lower their ask. The results for the regime dependent heteroskedasticity remain inconclusive.

Finally we discuss price discovery in an impulse response framework. We use a conditional bootstrap for the nonlinear impulse responses and find that the addition of nonlinearities in the model influence the outcome of these functions substantially. The initial situation of the model does not affect the outcome of the function, but the size of the shock does. According to the linear model, the last market maker dominates for Amgen and Instinet dominates for Intel. The inclusion of the nonlinearities reveal that for Amgen the first and the second market maker dominate. In case of Intel it is Island. Hence ignoring these nonlinearities would lead to different results.

## Chapter 4

# Price Discovery in Tick Time<sup>1</sup>

*The starting point of the analysis of price discovery is the specification of precisely what it is that is being discovered. In broad discussions of market efficiency, this is an almost theological question ...* – Lehmann (2002)

### 4.1 Introduction

Price discovery is the process of how different information sources contribute to the evolution of the underlying value of an asset. In a fragmented market with multiple dealers, like the Nasdaq, each dealer contributes to the price discovery process. Interesting questions are which dealer contributes most, how quick the discovery process works, and how it depends on market circumstances like liquidity, volatility and trading intensity. An important measure for the price discovery contribution of a dealer is the information share defined by Hasbrouck (1995). Huang (2002) provides an application to a variety of stocks on the Nasdaq to study the relationship between ECNs (Electronic Communication Networks) and traditional dealers.

Hasbrouck (1995) defines information shares as the part of the variance of the random walk component of returns that can be attributed to a particular market or dealer. But when quote updates of various dealers are contemporaneously correlated, the variance decomposition will not be unique. Hasbrouck (1995) suggests alternative Choleski decompositions to establish upper and lower bounds. For a particular stock at Nasdaq, Baillie, Booth, Tse, and Zobotina (2002) show, however, that upper and lower bounds can differ substantially, even if returns are observed at one minute intervals. When the sampling frequency is very high this contemporaneous correlation is minimal. For this reason, tick time data would be the preferred observation frequency. The use of tick time data for estimating price discovery among Nasdaq dealers is the main topic of this chapter.

---

<sup>1</sup>This Chapter is based on B. Frijns and P. Schotman, 'Price Discovery in Tick Time', CEPR Discussion Paper No. 4456.

This chapter provides three main contributions. First, we suggest a different model for the dynamics of quote changes. The traditional vector autoregressive of Hasbrouck (1995) is not suited for tick time data, as it always requires some time aggregation to construct a vector of quote updates for every time interval. An unobserved components specification is more suitable for ultra-high frequency data, since it does not require a complete vector of dealer quotes in each time interval. The model we propose is an extension of Hasbrouck (1993) to a setting with multiple dealers where quotes arrive in tick time.

Our second contribution is that we allow for duration effects on the quote dynamics. Time is an important factor in microstructure (Engle and Patton (2004)). It affects the volatility of the efficient price (Engle (2000)) and it also has an impact on the information content of dealer quote updates (Dufour and Engle (2000)). Both effects will be included in the model.

Our third contribution is the definition of measures for price discovery. These measures are extensions of the calendar time measures developed by De Jong and Schotman (2003) for an unobserved components model. We define these measures in tick time, where they are a function of the time between quote innovations. Additionally, we integrate these measures over time to reflect their calendar time equivalents. One of these calendar time measures resembles the Hasbrouck (1995) information shares. The structural interpretation of the unobserved components model provides a decomposition that leads to unambiguous information shares.

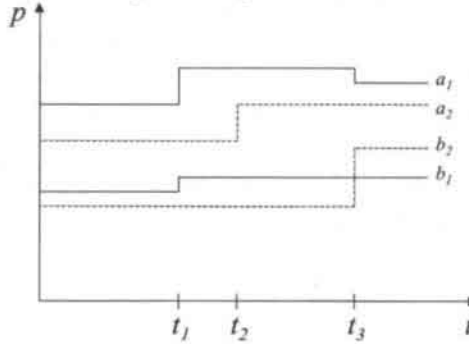
In our empirical part we examine the quotes of two ECN's (Island and Instinet) and three wholesale market makers at Nasdaq for 20 actively traded stocks with varying liquidity. These market makers are selected as the most active in terms of quoting frequency. For a tick time model it is more natural to consider the quote setting behavior of individual dealers instead of classes of dealers as in Huang (2002). Considering individual dealers is also in line with Schultz (2003) who finds a lot of heterogeneity among individual market makers.

As a preview of the results, we find that volatility does not evolve in calendar time, but either in tick time as mentioned by Clark (1973) or even less than tick time (see Dufour and Engle (2000)). The information flow to the efficient price is in general less at longer durations. We confirm the hypothesis of Easley and O'Hara (1992), which states that long durations convey no information. Similar results were found by Dufour and Engle (2000), and Engle and Patton (2004).

Price discovery measures in tick time appear strongly dependent on durations. Some dealers reveal information when durations are short whereas others reveal information when durations are long. Aggregating to calendar time we can often clearly identify the dominant dealer. In terms of price discovery we find that ECNs tend to dominate the liquid stocks, whereas market makers dominate the less liquid stocks. When the interval over which we aggregate increases, all these measures converge to single point estimates.

The remainder of this chapter is structured as follows. Section 4.2 discusses specification

Figure 4.1: Quote Arrivals



and estimation of the model. Section 4.3 defines the measures for price discovery in tick time and derives the calendar time aggregation of these measures. In section 4.4 we discuss the data. Section 4.5 presents the results of the model and the price discovery measures. Finally section 4.6 concludes.

## 4.2 A Model for Quotes in Tick Time

In this section we introduce a structural time series model for data in tick time and is an extension of the unobserved components model as proposed by Hasbrouck (1993). This model also fits the theoretical framework as proposed by Glosten and Harris (1988).

Consider a dealer market where  $M$  dealers issue bid and ask quotes. Quotes arrive at times  $t_\ell$  ( $\ell = 1, \dots, L$ ). The time between two consecutive quote arrivals is the quote duration. We are interested in modelling the dynamics of quote updates conditional on durations and assume that these durations are exogenous.

When sampling in tick time all quote changes are included. These can be changes in the bid and/or ask quotes of one or more dealers. In figure 4.1 we illustrate the quote arrival process. At  $t_1$  dealer 1 increases her bid and ask quote simultaneously. Next at  $t_2$ , dealer 2 increases only her ask. At time  $t_3$  both dealers change one of their quotes.

With  $M$  dealers in the market we have  $2M$  different time series of quote updates. Let  $q_\ell$  be the  $2M$ -vector of all standing quotes at time  $t_\ell$ . The bid (ask) of dealer  $i$  corresponds to element  $2i - 1$  ( $2i$ ) of  $q_\ell$ . To describe the quote dynamics in a model, we define  $I_\ell$  as the vector of identities of the quotes that have changed at  $t_\ell$ . This vector  $I_\ell$  has dimensions  $k_\ell \times 1$ . When, e.g. dealer  $i$  changes her bid at  $t_\ell$ ,  $I_\ell$  has a single element,  $I_\ell = 2i - 1$ . When dealer  $i$  changes both bid and ask,  $I_\ell = (2i - 1, 2i)$ . The vector of quotes that are updated at  $t_\ell$  can be written as  $q_{I_\ell, \ell}$ . The time series properties of updated quotes are modelled as



an unobserved components model. These components are a deterministic part ( $c$ ), which is a  $2M$ -vector due to the average spread between a quote series and the efficient price. This constant deviation captures e.g. the order processing costs the market maker faces. For a bid we expect negative elements in  $c$ , for an ask we expect these elements positive. The second part is a permanent part ( $m_\ell$ ), which is common to all dealers. The last part is a transitory part ( $u_\ell$ ), which is a  $2M$ -vector of factors not captured by  $c$  and  $m_\ell$ . This part captures the informational asymmetries among dealers, but also inventory costs and other kinds of noise. The decomposition for quotes follows as

$$q_{I_\ell, \ell} = c_{I_\ell} + \iota m_\ell + u_{I_\ell, \ell}, \quad (4.1)$$

where  $\iota$  is a unit vector.<sup>2</sup> The permanent part, also called the efficient price follows a random walk, with time dependent volatility ( $\sigma_\ell$ ). The impact of time on the volatility of the price process was questioned by Clark (1973), Harris (1987) and Ané and Geman (2000), among others.

$$m_\ell = m_{\ell-1} + \sigma_\ell r_\ell. \quad (4.2)$$

The innovation term  $r_\ell$  has unit variance. We specify volatility as

$$\sigma_\ell = \sigma \tau_\ell^{\delta_1}, \quad (4.3)$$

where  $\tau_\ell$  is the duration between quotes, normalized by dividing it by the average quote duration over the whole sample.<sup>3</sup> The parameter  $\delta_1$  measures the impact of quote durations on the volatility of the random walk. If  $\delta_1 = \frac{1}{2}$  the random walk is said to evolve in calendar time. In this case the variance of  $(m_p - m_s)$  is equal to  $(\sum_{\ell=s+1}^p \tau_\ell) \sigma^2$ , and thus proportional to the length of the calendar time interval. When  $\delta_1 = 0$  then the variance of the random walk is not affected by the time between quote updates. In this case the random walk evolves in tick time. The calendar time variance is proportional to the number of quote updates (see e.g. Clark (1973)).

The last term in (4.1) is the transitory term  $u_\ell$ , which measures the temporary deviations from the efficient price and includes asymmetric information, inventory effects and other sources of noise. To capture the asymmetric information we allow the transitory component for every dealer, to correlate with the innovation in the efficient price. This is in line with the theoretical framework of Glosten and Harris (1988), who argue that the prices of informed market participants are correlated with the innovation in the efficient price. The

<sup>2</sup>The length of this unit vector is suppressed at the length of the vector of quote updates.

<sup>3</sup>This normalization only affects the  $\sigma$  parameter, which now refers to the volatility per tick, instead of the volatility per second.

decomposition for the transitory component therefore reads

$$u_{I_\ell, \ell} = \alpha I_\ell \tau_\ell^{\delta_2} \sigma r_\ell + e_{I_\ell, \ell}. \quad (4.4)$$

where the asymmetric information is captured by  $\alpha$ . We allow this information component to depend on the duration between quotes. The impact time has on the asymmetric information is measured by  $\delta_2$ . In case  $\delta_2 > 0$ , quotes become more informative at longer durations. If  $\delta_2 < 0$  the opposite holds. The idiosyncratic quote noise ( $e_{I_\ell, \ell}$ ) is uncorrelated with  $r_\ell$  and  $\tau_\ell$  and has covariance matrix  $\Omega$ .

To estimate the model we put it in state space form for a time series process with missing observations. Following Harvey (1989) we write

$$\begin{aligned} J'_\ell q_\ell &= J'_\ell c + J'_\ell m_\ell + J'_\ell u_\ell, \\ m_\ell &= m_{\ell-1} + \sigma_\ell r_\ell, \\ J'_\ell u_\ell &= J'_\ell \alpha \tau_\ell^{\delta_2} \sigma r_\ell + J'_\ell e_\ell, \end{aligned} \quad (4.5)$$

where  $J_\ell$  is a matrix that selects the elements from the quote vector that are updated. This matrix has dimension  $2M \times k_\ell$ . In this case  $J'_\ell q_\ell$  is a  $k_\ell \times 1$  vector of updated quotes. This pre-multiplication is applied to all the components that we discussed.

An advantage of putting the model in a state space form is that with some additional restrictions the parameters  $c$ ,  $\alpha$ ,  $\delta_1$ ,  $\delta_2$ ,  $\sigma$  and  $\Omega$  can be identified. The first restriction is bid and ask deviations from the efficient price are symmetric. This restriction is merely imposed to reduce the number of parameters in the model. The second restriction is on the idiosyncratic quote noise ( $e_\ell$ ). We assume the covariance matrix of this noise term ( $\Omega$ ) to have a block diagonal structure. This structure only allows bid and ask quotes of the same dealer to be correlated. We motivate this structure by the fact that the quote noise includes inventory effects and a remaining microstructure noise. Theoretical models for inventory costs adhere to the notion that when a dealer receives inventory she will lower her ask to induce a trade at the opposite side, but also lowers her bid to avoid receiving additional inventory. Therefore, a specific dealer may wish to alter both quotes simultaneously due to the inventory position she has. Simultaneous changes in the quotes of different dealers can be due to asymmetric information both dealers have, which is captured by  $\alpha$  or occurs by mere chance.

De Jong and Schotman (2003) show that for a model in calendar time, imposing structure on  $\Omega$  is necessary for the identification of  $\alpha$ . In tick time there are a few additional effects that hinder this identification. When the dimension of the quote vector is one at each observation, the identification of  $\alpha$  is not guaranteed. Without duration functions  $\alpha$  can only be identified up to a translation along the unit vector. See Appendix 4.6 for a full derivation. Duration

functions help to make identification possible, but only if  $\delta_1 \neq 0$  and  $\delta_2 \neq 0$ . When these measures are close to zero identification will be troublesome.

Estimation of (4.5) is done using the Kalman Filter. As the underlying process is a random walk, the system cannot be initialized using the long run mean of the underlying process. We use a diffuse prior to initialize the random walk process. To incorporate the potentially large price change overnight the same approach is followed. Everyday the model is re-initialized with a diffuse prior. The first 50 observations on each day are excluded in the maximization of the Likelihood function as the variance of the prediction error has not converged to normal levels.

As the model considers quote updates, the dimension of the quote vector will frequently be one. This makes computations in the Kalman Filter recursions straightforward as determinants and inverses are given. However since this dimension changes over time the model will not converge to a steady state and the full recursion will have to be computed for each observation. For this reason the filtering routine is computationally demanding.

## 4.3 Price Discovery

In this section we define price discovery measures in tick time using the unobserved components model of the previous section. These tick time measures are subsequently aggregated to calendar time equivalents.

### 4.3.1 Price Discovery in Tick time

We consider three measures to summarize the quote setting behavior of dealers. The first measure considers how dealers incorporate information in the efficient price into their quote innovations. The second measure considers the contribution of quote innovations to the evolution of the efficient price. The last measure combines both measures in one single measure, which considers the total information share of dealers.

To explore the implications of the model we consider the following scenario. Suppose at  $t_\ell$  each dealer issues bid and ask quotes, and suppose that the previous efficient price  $m_{\ell-1}$  is known to all dealers. The quote updates reflect both the change in efficient price,  $m_\ell - m_{\ell-1}$  between  $t_{\ell-1}$  and  $t_\ell$  and the dealer noise  $e_\ell$ . The innovation of the dealer quotes is then equal to

$$\begin{aligned} v_\ell &= q_\ell - E_{\ell-1}[q_\ell] = \iota(m_\ell - m_{\ell-1}) + u_\ell \\ &= (\iota + \alpha\tau_\ell^{\delta_2 - \delta_1})(m_\ell - m_{\ell-1}) + e_\ell. \end{aligned} \quad (4.6)$$

The amount of information in the efficient price that is incorporated into quotes is referred to as *dealer efficiency*. The more of the innovation in the random walk that is incorporated

in dealer quotes, the more efficient they are. To obtain a measure for dealer efficiency we consider the covariance between quote updates as a fraction of the total variance of the random walk

$$\beta(\tau) = \text{Cov}(v_\ell \Delta m_\ell) / \text{Var}(\Delta m_\ell) = \iota + \alpha \tau^{(\delta_2 - \delta_1)}, \quad (4.7)$$

which is a function of the duration of quote innovations. The crucial parameters are  $\alpha$ ,  $\delta_1$  and  $\delta_2$ . At long durations this measure will converge to 1 when  $\delta_1 > \delta_2$ , and diverge in the opposite case. The sign of  $\alpha$  determines whether 1 is an upper or a lower bound for the respective element in  $\beta(\tau)$ . When  $\alpha$  is larger than zero,  $\beta(\tau)$  will be larger than one. This indicates that this quote incorporates more information of the random walk and vice versa. Hence whether a dealer is more or less efficient depends solely on the sign of  $\alpha$ . Whether this efficiency persists over long or short durations is dependent on the values of  $\delta_1$  and  $\delta_2$ .

For the second measure we consider a variance decomposition of the efficient price similar to Hasbrouck's (1995) analysis of *price discovery*. It relates the change in the efficient price to innovations in dealer quotes,

$$\Delta m_\ell = \gamma(\tau)' v_\ell + \varepsilon_\ell, \quad (4.8)$$

where the regression coefficients  $\gamma(\tau)$  are defined as

$$\gamma(\tau) = \text{Var}(v_\ell)^{-1} \text{Cov}(v_\ell \Delta m_\ell). \quad (4.9)$$

For the covariance matrix  $\Sigma(\tau) = \text{Var}(v_\ell)$  we find

$$\Sigma(\tau) = \Omega + \sigma^2 \tau^{2\delta_1} (\iota + \alpha \tau^{\delta_2 - \delta_1})(\iota + \alpha \tau^{\delta_2 - \delta_1})'. \quad (4.10)$$

The covariance follows as

$$\text{Cov}(v_\ell \Delta m_\ell) = \sigma^2 \tau^{2\delta_1} (\iota + \alpha \tau^{\delta_2 - \delta_1}). \quad (4.11)$$

To arrive at an expression for  $\gamma(\tau)$  we apply the matrix inversion lemma on  $\Sigma(\tau)$ ,

$$\Sigma(\tau)^{-1} = \Omega^{-1} - \frac{\sigma^2 \tau^{2\delta_1} \Omega^{-1} (\iota + \alpha \tau^{\delta_2 - \delta_1})(\iota + \alpha \tau^{\delta_2 - \delta_1})' \Omega^{-1}}{\sigma^2 \tau^{2\delta_1} (\iota + \alpha \tau^{\delta_2 - \delta_1})' \Omega^{-1} (\iota + \alpha \tau^{\delta_2 - \delta_1})}, \quad (4.12)$$

and therefore find

$$\gamma(\tau) = \frac{\sigma^2 \tau^{2\delta_1} \Omega^{-1} (\iota + \alpha \tau^{\delta_2 - \delta_1})}{1 + \sigma^2 \tau^{2\delta_1} (\iota + \alpha \tau^{\delta_2 - \delta_1})' \Omega^{-1} (\iota + \alpha \tau^{\delta_2 - \delta_1})}. \quad (4.13)$$

This expression of  $\gamma(\tau)$  can also be written as a function of  $\beta(\tau)$

$$\gamma(\tau) = \frac{\sigma^2 \tau^{2\delta_1} \Omega^{-1} \beta(\tau)}{1 + \sigma^2 \tau^{2\delta_1} \beta(\tau)' \Omega^{-1} \beta(\tau)}. \quad (4.14)$$

Hence this measure combines the dealer efficiency measure and the amount of idiosyncratic noise each dealer quote has.

The last measure resembles the information shares of Hasbrouck (1995). These are defined as the proportional contribution of dealer quote innovations to the innovation in the efficient price. This is determined by considering how much of the variance of the efficient price can be explained by the variance in quote innovations. This can be expressed in terms of the  $R^2$  of regression (4.8). From standard regression theory it is known that this is the inner product of  $\gamma(\tau)$  and  $\beta(\tau)$

$$R^2(\tau) = \gamma(\tau)' \beta(\tau). \quad (4.15)$$

Hasbrouck (1995) information shares have the characteristic that they cannot be assigned to dealers uniquely. Given that the matrix inversion lemma can be applied and the block diagonal structure of  $\Omega$ ,  $R^2(\tau)$  can be decomposed as the sum of the information shares per dealer ( $IS_i(\tau)$ ).

### 4.3.2 Calendar Time Aggregation

In the previous section we discussed price discovery and efficiency of single quote updates. This section defines similar measures of quote setting behavior over time. To obtain these measures we aggregate the tick time measures to fixed time intervals. Let us first establish the relationship between tick time and calendar time. In calendar time we refer to  $t$  as the present time and  $t + 1$  as the time at the next time interval. Let  $Q(t)$  be the vector of quotes at time  $t$ . Assume that deviations from the efficient price are included in the information set of all dealers. In calendar time innovations in dealer quotes are decomposed as

$$V(t) = Q(t) - E_{t-1}[Q(t)] = \iota(m(t) - m(t-1)) + u(t). \quad (4.16)$$

The innovation in quotes is equal to the change in the efficient price plus the noise around this efficient price. To establish the link between calendar time and tick time, let  $\ell_t$  represent the observation closest to time  $t$ . The tick time equivalent (4.16) is

$$v_{\ell_t} = \iota(m_{\ell_t} - m_{\ell_{t-1}}) + u_{\ell_t}. \quad (4.17)$$

The change in the efficient price from  $\ell_{t-1}$  to  $\ell_t$  is the sum of all interjacent changes. We write (4.17) as the sum of these changes

$$v_{\ell_t} = \iota(m_{\ell_t} - m_{\ell_{t-1}}) + u_{\ell_t} = \iota \sum_{j=\ell_{t-1}+1}^{\ell_t} (m_j - m_{j-1}) + u_{\ell_t}. \quad (4.18)$$

Dealer efficiency is considered as the change in the dealer quotes that can be attributed to the change in the efficient price. This is the regression coefficient of quote innovations regressed on the change in the efficient price. We obtain this measure by considering the covariance of quote innovations with the change in the efficient price as a fraction of the total variance caused by the efficient price. The covariance follows as

$$\begin{aligned} Cov(v_{\ell_t}, (m_{\ell_t} - m_{\ell_{t-1}})) &= E[(\iota \sum_{j=\ell_{t-1}+1}^{\ell_t} \tau_j^{\delta_1} \sigma r_j + \alpha \tau_{\ell_t}^{\delta_2} \sigma r_{\ell_t}) (\sum_{j=\ell_{t-1}+1}^{\ell_t} \tau_j^{\delta_1} \sigma r_j)] \\ &= \sigma^2 (\iota \sum_{j=\ell_{t-1}+1}^{\ell_t} \tau_j^{2\delta_1} + \alpha \tau_{\ell_t}^{\delta_1+\delta_2}). \end{aligned} \quad (4.19)$$

The variance of the random walk is

$$Var(m_{\ell_t} - m_{\ell_{t-1}}) = \sigma^2 \sum_{j=\ell_{t-1}+1}^{\ell_t} \tau_j^{2\delta_1} \quad (4.20)$$

The measure for dealer efficiency follows from this,

$$\beta(t) = \iota + \alpha \frac{\tau_{\ell_t}^{\delta_1+\delta_2}}{\sum_{j=\ell_{t-1}+1}^{\ell_t} \tau_j^{2\delta_1}}. \quad (4.21)$$

Again  $\beta(t)$  is a re-scaling of  $\alpha$ . As the interval between  $\ell_t$  and  $\ell_{t-1}$  increases,  $\beta(t)$  converges to unity. Thus when intervals are larger, dealers incorporate information about the true price more efficiently into their quotes. The rate of convergence is dependent on  $\alpha$ ,  $\delta_1$  and  $\delta_2$ . Depending on the sign of  $\alpha$  this convergence will occur from above ( $\alpha > 0$ ) or below ( $\alpha < 0$ ). When convergence in  $\beta(t)$  occurs quickly then dealers quote efficiently in this stock.

For price discovery we consider the reverse regression as for dealer efficiency. We consider the change in the efficient price that can be attributed to innovations in dealer quotes

$$\gamma(t)' = Cov(V_{\ell_t}, (m_{\ell_t} - m_{\ell_{t-1}})) \Sigma(t)^{-1}. \quad (4.22)$$

The variance of quote innovations follows as

$$\begin{aligned}\Sigma(t) &= \text{Var}(\iota(m_{\ell_t} - m_{\ell_{t-1}} + u_{\ell_t})) \\ &= \sigma^2(\iota' \sum_{\ell=\ell_{t-1}+1}^{\ell_t-1} \tau_{\ell}^{2\delta_1} + (\iota\tau_{\ell_t}^{\delta_1} + \alpha\tau_{\ell_t}^{\delta_2})(\iota\tau_{\ell_t}^{\delta_1} + \alpha\tau_{\ell_t}^{\delta_2})') + \Omega.\end{aligned}\quad (4.23)$$

This matrix depends on all the interjacent durations from  $\ell_{t-1}$  to  $\ell_t$ . Therefore the inverse of this matrix is not straightforward. To extract the duration dependence, the matrix inversion lemma needs to be applied twice. For the sake of notational comfort the inverse is not displayed.

As mentioned before information shares are determined by the  $R^2$  of the regression of dealer quotes on the efficient price. Again this measure is the inner product of dealer efficiency and price discovery

$$R^2(t) = \gamma(t)' \beta(t). \quad (4.24)$$

Here the inversion lemma does not only extract the durations from the calculation of the inverse. For information shares this lemma is needed to assign them to dealers individually. As the inverse of  $\Sigma(t)$  is a function of the inverse of  $\Omega$ , which was known to have a block diagonal structure, these information shares can be assigned uniquely.

When the sampling interval increases the change in dealer quotes will to a large extent represent the change in the efficient price. As a consequence  $R^2(t)$  converges to one. This convergence of  $R^2(t)$  will be related to the convergence of  $\beta(t)$  to a unit vector. Although stated without proof, if  $R^2(t)$  and  $\beta(t)$  converge to unity, then the sum of the elements in  $\gamma(t)$  add up to one. As the covariance between quote innovations and the change in the random walk is positive, as well as the variance of the random walk, elements in  $\gamma(t)$ , will converge to single point estimates between zero and one.

The total value of  $R^2(t)$  indicates how much microstructure noise there is left in dealer quotes. Results from this could have consequences for e.g. the sampling intervals used for realized volatility as described by Andersen, Bollerslev, Diebold, and Labys (2001). When  $R^2(t)$  is low there is still a lot of microstructure noise in the data and longer sampling intervals should be used.

## 4.4 Data

The data used in this study is provided by Nastroq. This data set contains all trades and quotes that occur within normal trading hours at Nasdaq. From this data set we consider dealer quote data. This provides all quotes issued within trading hours, time stamped to the nearest second. Most important, it contains the identity of the dealer that issues the quote.

Table 4.1: List of Company Ticker symbols and Company names

Ticker Symbol	Company name
AAPL	Apple Computer Inc.
AMAT	Applied Materials Inc.
AMGN	Amgen Inc.
AMZN	Amazon.com, Inc.
ATHM	At Home Corporation
CMGI	CMGI, Inc.
COMS	3Com Corporation
CPWR	Compuware Corporation
CSCO	Cisco Systems Inc.
DELL	Dell Computer Corporation
INTC	Intel Corporation
MSFT	Microsoft Corporation
NOVL	Novell Inc.
NXTL	Nextel Communications CL-A
ORCL	Oracle Corporation
PSFT	Peoplesoft Inc.
QWST	Qwest Communications Intl Inc.
SBUX	Starbucks Corporation
SUNW	Sun Microsystems Inc.
WCOM	MCI WorldCom Inc.

For our study we select 20 highly traded companies with different liquidity listed at Nasdaq for February 1999. The selected stock and their ticker symbols are reported in table 4.1. Huang (2002) uses the same data set as ours in 1999 but uses different months in his study. He creates categories of different types of dealers. Schultz (2003) argues that dealer quoting behavior is heterogeneously, hence we consider individual dealers. We consider the dealer quotes of the 5 largest dealers, in terms of quoting frequency. This leads to the selection of 2 ECNs, Island and Instinet, and 3 market makers, which change depending on the stock considered.

The data provided needs to be filtered before it can be used in the model. As we consider the innovation in a quote (either bid or ask quote) the other quotes are removed from the data. First, when multiple quotes are issued at the same second, the last quote in this sequence is selected.<sup>4</sup> Second, a change in the depth of the quote is not considered. All quotes that do not change the bid or ask are removed. When a dealer only updates a bid this is the only quote that we will consider in the model. Due to this selection criteria the number of newly issued quotes is often 1.

Another issue is the treatment of outliers. When a dealer is unwilling to trade she will issue a quote far away from the inside (best quote in the market). This can happen regularly on one side of the market when a dealer is unwilling to take more inventory. Although these

<sup>4</sup>Especially in the case of Island this occurs frequently. This is caused since there are many small trades matched within Island itself. Every time this happens Island will send the new best standing quote to Nasdaq, which in many cases is merely a change in the depth of the quote



Table 4.2: Summary statistics

Stock	# of Obs	% of single quotes	Average Duration	Std of Duration
AAPL	29,787	89.63	15.71	28.73
AMAT	105,090	83.58	4.77	5.95
AMGN	40,279	83.97	12.32	22.24
AMZN	150,710	80.52	3.44	4.39
ATHM	76,435	86.50	6.34	9.40
CMGI	90,401	87.06	5.34	7.75
COMS	61,049	89.56	7.68	11.34
CPWR	33,301	90.55	13.92	31.03
CSCO	164,480	80.85	3.13	3.52
DELL	177,850	77.39	3.02	3.44
INTC	171,260	76.92	3.15	3.56
MSFT	151,110	80.82	3.42	3.94
NOVL	18,909	87.88	25.08	43.93
NXTL	19,556	91.63	23.42	42.23
ORCL	87,774	85.81	5.56	7.46
PSFT	24,601	91.12	18.74	32.02
QWST	44,459	88.331	10.68	18.12
SBUX	14,320	90.056	32.43	64.92
SUNW	128,370	82.43	4.20	5.15
WCOM	88,550	83.33	5.66	7.41

*Note:* This table presents some summary statistics of the data after being filtered. Column 1 represents the total number of quotes issued for each stock. Column 2 indicates in how many cases there is only 1 quote innovation in the observation matrix. Column 3 gives the average duration between any quote updates and column 4 gives the respective standard deviation.

quotes send a very strong signal to the other market participants, this is something not considered here as we address the issues of price discovery. We define an outlier as a quote that is more than \$ 2 away from the average of the past 50 quotes. These outliers are deleted.

Table 4.2 reports some summary statistics of the data after filtering. The first column reports the total number of quotes given by dealers in each stock. Our data set incorporates very liquid stocks like DELL as well as illiquid stocks like SBUX. The second column report the fraction of single quote updates. In section 4.2 we discussed the identification problems that could occur. The number of quotes is inversely related to the duration between quotes. Further note that there is a weak relationship between the percentage of single quotes and the total number of quotes issued. The other issue that can lead to identification problems is the variance in durations. These seem to be large enough not to cause any identification problems.

Table 4.3: Duration Parameters

Stock	$\delta_1$	$\delta_2$
AAPL	-0.06 (0.03)	-0.44 (0.05)
AMAT	0.02 (0.02)	-1.20 (0.17)
AMGN	0.04 (0.02)	-0.21 (0.17)
AMZN	0.03 (0.01)	0.50 (0.02)
ATHM	0.01 (0.02)	-1.26 (0.18)
CMGI	0.07 (0.01)	-0.11 (0.04)
COMS	-0.11 (0.02)	-0.51 (0.06)
CPWR	0.00 (0.02)	-0.72 (0.13)
CSCO	-0.07 (0.01)	-0.24 (0.04)
DELL	-0.13 (0.01)	-0.30 (0.03)
INTC	-0.17 (0.02)	-0.23 (0.04)
MSFT	-0.20 (0.02)	-0.13 (0.03)
NOVL	-0.23 (0.04)	-0.59 (0.05)
NXTL	-0.05 (0.03)	-0.44 (0.08)
ORCL	0.02 (0.02)	-0.82 (0.05)
PSFT	0.00 (0.02)	-0.23 (0.03)
QWST	-0.14 (0.02)	0.23 (0.03)
SBUX	-0.01 (0.03)	-0.32 (0.06)
SUNW	-0.06 (0.02)	-0.96 (0.11)
WCOM	-0.11 (0.02)	-1.17 (0.14)

*Note:* This table reports the estimates and standard errors of the duration parameters in the model,

$$\begin{aligned}
 J'_\ell q_\ell &= J'_\ell c + J'_\ell \epsilon m_\ell + J'_\ell u_\ell, \\
 m_\ell &= m_{\ell-1} + \tau_\ell^{\delta_1} r_\ell, \\
 J'_\ell u_\ell &= J'_\ell \alpha \tau_\ell^{\delta_2} r_\ell + J'_\ell e_\ell,
 \end{aligned}$$

Column 1 gives the duration parameters on the random walk component, column 2 reports the duration parameters of the correlation between dealer quotes and the random walk.

## 4.5 Results

This section provides the results of the model in section 4.2 and the measures of price discovery discussed in section 4.3.

### 4.5.1 Parameter Estimates

In this part we discuss the parameter estimates from model (4.5). In table 4.3 we report the estimates for the duration parameters, including their standard errors. The parameter  $\delta_1$  measures the duration effect on the innovation in the efficient price. This parameter is significantly negative in 50% of the cases or zero in most other cases. For most stocks the random walk evolves in tick time or we find that long durations between quotes decreases the volatility of the random walk, which is consistent with the findings of Dufour and Engle

Table 4.4: Quote Specific Relations to the Efficient Price

Stock	ISLD <sub>b</sub>	ISLD <sub>a</sub>	INCA <sub>b</sub>	INCA <sub>a</sub>	MM1 <sub>b</sub>	MM1 <sub>a</sub>	MM2 <sub>b</sub>	MM2 <sub>a</sub>	MM3 <sub>b</sub>	MM3 <sub>a</sub>
AAPL	-1.02 (0.20)	-0.56 (0.21)	1.24 (0.37)	0.58 (0.26)	3.17 (0.40)	2.01 (0.35)	0.98 (0.19)	-1.49 (0.25)	-0.82 (0.36)	-0.21 (0.28)
AMAT	-0.03 (0.02)	-0.06 (0.02)	0.07 (0.05)	0.38 (0.10)	0.49 (0.13)	0.33 (0.11)	-0.08 (0.05)	-0.12 (0.06)	0.02 (0.14)	-0.63 (0.19)
AMGN	-1.59 (0.43)	-0.69 (0.42)	0.13 (0.20)	0.14 (0.28)	-1.44 (0.69)	-0.97 (0.60)	-1.70 (0.93)	-1.70 (1.09)	0.17 (0.41)	-0.13 (0.43)
AMZN	0.00 (0.05)	0.12 (0.05)	0.93 (0.08)	1.38 (0.08)	-0.33 (0.12)	-0.66 (0.13)	-0.97 (0.16)	-5.21 (0.13)	-0.12 (0.20)	-0.05 (0.18)
ATHM	-0.08 (0.03)	-0.11 (0.04)	0.08 (0.04)	0.04 (0.03)	0.38 (0.13)	0.28 (0.10)	-0.15 (0.06)	-0.16 (0.07)	-0.13 (0.08)	-0.17 (0.09)
CMGI	-0.35 (0.08)	-0.61 (0.09)	0.35 (0.15)	0.44 (0.18)	-0.39 (0.17)	-0.11 (0.19)	1.68 (0.15)	1.39 (0.15)	-1.03 (0.21)	-1.46 (0.23)
COMS	-0.07 (0.08)	-0.02 (0.07)	1.27 (0.17)	0.79 (0.18)	-1.21 (0.28)	-1.02 (0.30)	0.00 (0.14)	-0.24 (0.22)	0.00 (0.17)	1.49 (0.28)
CPWR	-0.22 (0.09)	-0.31 (0.11)	-0.12 (0.08)	0.04 (0.08)	0.59 (0.21)	0.37 (0.19)	0.79 (0.32)	0.89 (0.36)	0.12 (0.18)	-0.26 (0.18)
CSCO	-0.41 (0.07)	-0.42 (0.08)	0.44 (0.10)	0.48 (0.09)	0.55 (0.18)	0.03 (0.25)	0.76 (0.20)	0.80 (0.25)	-3.00 (0.22)	-2.92 (0.23)
DELL	-0.47 (0.04)	-0.47 (0.04)	0.12 (0.08)	0.26 (0.07)	4.54 (0.34)	5.34 (0.37)	0.01 (0.12)	0.60 (0.16)	-2.39 (0.11)	-2.84 (0.18)
INTC	0.00 (0.09)	-0.35 (0.08)	-0.28 (0.09)	-0.03 (0.08)	0.67 (0.15)	1.04 (0.16)	-1.39 (0.10)	-1.01 (0.11)	-1.39 (0.20)	-1.55 (0.19)
MSFT	-0.07 (0.09)	-0.07 (0.09)	-0.10 (0.10)	0.05 (0.09)	-0.81 (0.19)	-0.18 (0.23)	-1.31 (0.12)	-1.22 (0.18)	-2.80 (0.17)	-2.73 (0.18)
NOVL	1.16 (0.32)	-0.07 (0.12)	1.69 (0.28)	0.16 (0.11)	1.73 (0.32)	-2.58 (0.44)	0.30 (0.15)	1.00 (0.20)	0.15 (0.16)	-0.07 (0.16)
NXTL	-1.44 (0.39)	-1.25 (0.42)	-0.49 (0.23)	-0.01 (0.34)	1.89 (0.45)	1.23 (0.42)	1.51 (0.48)	-1.32 (0.62)	0.16 (0.20)	-1.50 (0.38)
ORCL	-0.34 (0.06)	-0.46 (0.06)	0.20 (0.07)	-0.04 (0.09)	2.62 (0.22)	-2.15 (0.18)	-0.27 (0.07)	-0.05 (0.08)	-0.42 (0.12)	-0.26 (0.15)
PSFT	0.93 (0.36)	-0.43 (0.21)	1.90 (0.30)	0.18 (0.26)	11.98 (1.44)	9.51 (1.28)	8.09 (1.03)	9.20 (1.14)	-0.24 (0.54)	2.33 (0.56)
QWST	0.99 (0.69)	0.22 (0.43)	4.57 (0.47)	5.46 (0.42)	-0.45 (0.42)	0.23 (0.63)	2.79 (0.71)	3.44 (1.11)	0.70 (0.19)	0.37 (0.21)
SBUX	0.14 (0.19)	-0.08 (0.32)	1.21 (0.34)	0.22 (0.23)	1.87 (0.41)	1.28 (0.31)	-0.52 (0.34)	-1.10 (0.43)	-0.27 (0.24)	-0.16 (0.19)
SUNW	-0.08 (0.06)	-0.31 (0.08)	-0.04 (0.07)	0.00 (0.06)	-0.55 (0.11)	-0.50 (0.14)	-1.07 (0.19)	-1.61 (0.27)	0.10 (0.11)	0.10 (0.13)
WCOM	-0.02 (0.04)	-0.12 (0.05)	0.24 (0.06)	-0.13 (0.05)	0.65 (0.17)	0.49 (0.15)	-0.26 (0.08)	-0.36 (0.09)	0.17 (0.08)	-0.05 (0.06)

Note: This table presents the estimates for  $\alpha$  and their standard errors. These  $\alpha$ 's are estimated in

$$J'_\ell q_\ell = J'_\ell c + J'_\ell m_\ell + J'_\ell u_\ell,$$

$$m_\ell = m_{\ell-1} + \tau_\ell^{\delta_1} r_\ell,$$

$$J'_\ell u_\ell = J'_\ell \alpha \tau_\ell^{\delta_2} r_\ell + J'_\ell e_\ell.$$

The subscripts  $b$  and  $a$  represent the respective bid or ask side of a dealer.

(2000). The value of  $\delta_1 = \frac{1}{2}$ , which represents a random walk in calendar time can easily be rejected.

The parameter  $\delta_2$  measures the duration dependence of the asymmetric information. In most cases this parameter is negative and more negative than  $\delta_1$ . Longer durations between

quote updates result in less asymmetric information about the efficient price. Hence at short durations quotes are more informative. These results are similar to the findings of Dufour and Engle (2000) and Engle and Patton (2004). They find that the price impact of trades is larger at short durations. Our results also confirm the reasoning of Easley and O'Hara (1992). Their model hypothesizes that long durations convey no information. The combined results of the duration parameters indicate that short durations means higher volatility, and more asymmetric information. As a consequence volatile periods are periods where asymmetries are large and vice versa.

In table 4.4 we report the  $\alpha$ 's from (4.5) including their standard errors. An  $\alpha$  significantly larger/smaller than zero means that this dealer has more/less exposure the innovation in the efficient price than the average of dealers. Although these  $\alpha$ 's are not that informative on their own, they do indicate large heterogeneity among the different dealers. Moreover it also indicates the heterogeneity present in the bid and ask quotes themselves. This heterogeneity among stocks will cause different results when discussing price discovery. It also indicates that dealer cannot just be grouped in categories as their characteristics are very different among individual dealers.

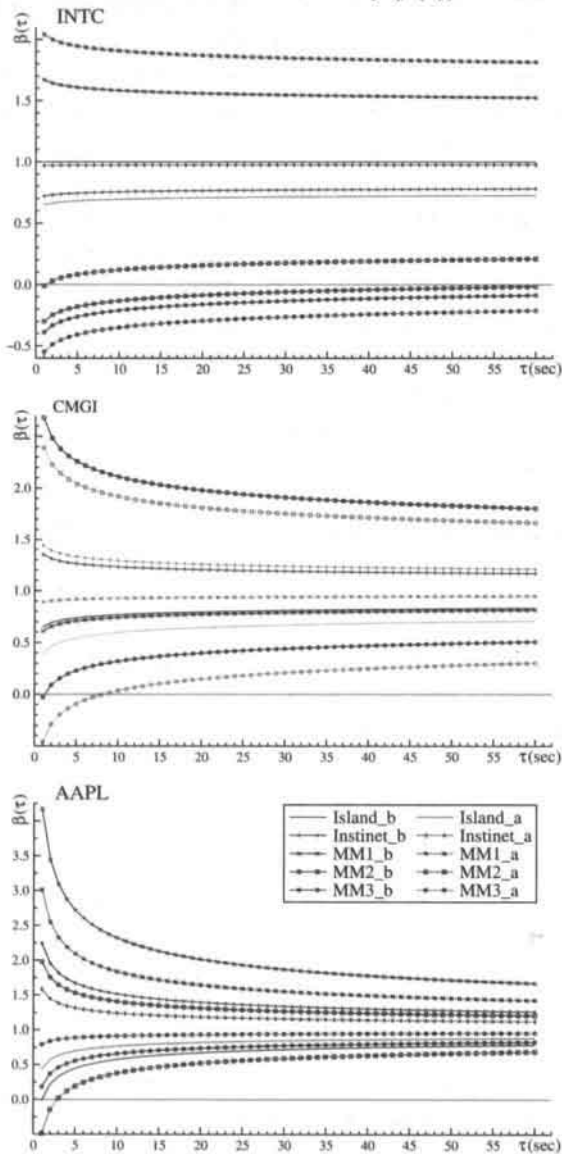
#### 4.5.2 Price Discovery in Quote Updates

In this section we discuss the results of these measure per quote update.

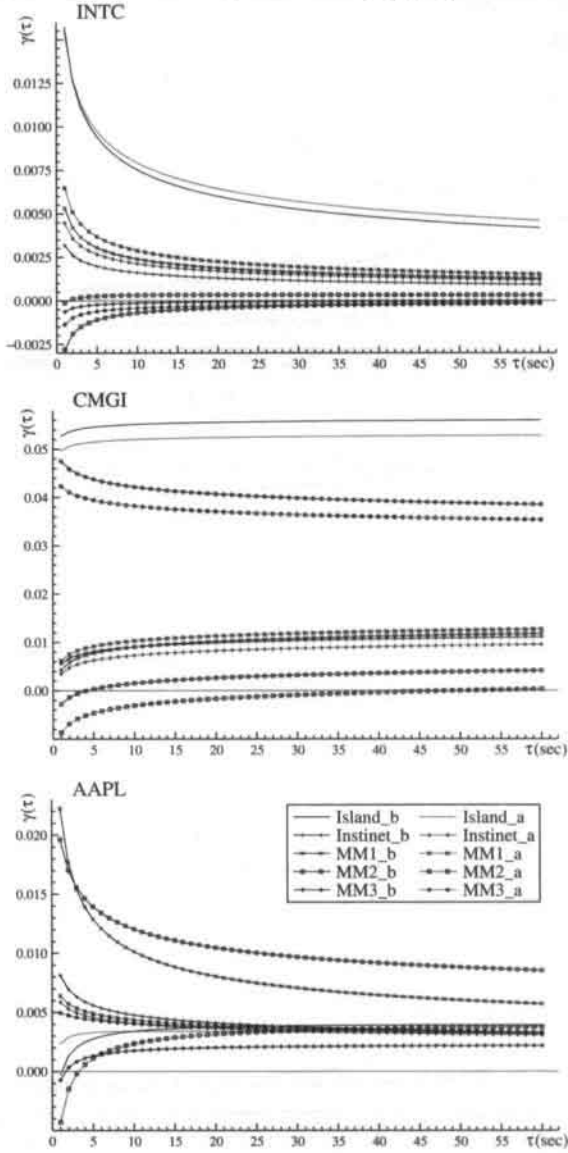
To discuss dealer efficiency we consider the innovations in quotes that are due to a change in the efficient price. This measure is a re-scaling of the  $\alpha$ 's reported before, and is dependent on the duration estimates for their convergence. In most of the cases  $\delta_1$  is larger than  $\delta_2$ , which means that for long durations the estimate for  $\beta(\tau)$  will converge to one. This is expected, as when there is a long time between quote updates, the change in quotes is likely to be caused by a change in the efficient price. As the measures for  $\beta(\tau)$  are duration dependent, they cannot be shown in a table. In figure 4.2 we draw  $\beta(\tau)$  for three specific stocks in the sample as a function of duration. These stocks were selected to represent liquid stocks (INTC), less liquid stocks (CMGI) and illiquid stocks (AAPL). For all stocks  $\beta(\tau)$  converges to one, as  $\delta_2 < \delta_1$ . Even when durations are as long as one minute, the change in quotes can still not be explained fully by the change in the underlying price. As there was heterogeneity among the  $\alpha$ 's of the different stock there is the same heterogeneity in  $\beta(\tau)$ .

For measures of price discovery in tick time we perform the reverse regression of the efficient price on quote innovations. We consider to what extent a single innovation in a dealer quote can explain the change in the efficient price. Again this measure is duration dependent and should be considered relatively with respect to Island. In figure 4.3 we show the same stocks as before. The dynamics of these estimates cannot be derived straightforward from the parameters as for  $\beta(\tau)$ . In the case of INTC we see a clear dominance of the ECNs

Figure 4.2: Tick time measure for dealer efficiency ( $\beta(\tau)$ ) as a function of duration

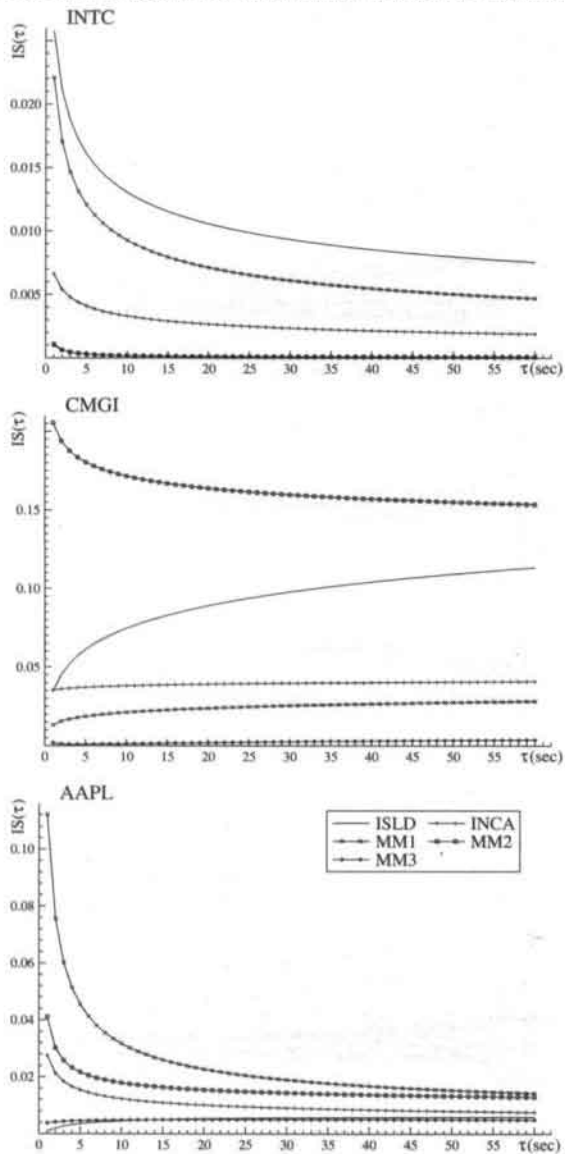


Note: These graphs show the dealer efficiency measures in tick time for INTL, CMGI and AAPL. These measures are plotted as a function of duration.

Figure 4.3: Tick time measure for price discovery ( $\gamma(\tau)$ ) as a function of duration

*Note:* These graphs show the price discovery measures in tick time for INTC, CMGI and AAPL. These measures are plotted as a function of duration.

Figure 4.4: Tick time measure for Information Shares as a function of duration



Note: These graphs show information shares in tick time for INTC, CMGI and AAPL per dealer. These measures are plotted as a function of duration.

over the market makers. The price discovery is highest at very short durations, and declines rapidly when durations get longer. However for INTC durations most often are very short. For CMGI, Island dominates persistently. The price discovery for Island does not decrease as the durations increase. An explanation for this is that the  $\alpha$ 's for Island are close to zero and the duration parameter  $\delta_2$  is also close to zero. For AAPL we see a similar pattern as for INTC. Interesting is the crossing line for MM1 and MM2 at very short durations.

The information shares we consider are determined by the  $R^2$  of the regression of the change in the efficient price on the innovation in dealer quotes. This is the inner product of  $\beta(\tau)$  and  $\gamma(\tau)$ . This  $R^2(\tau)$  is decomposed to dealer specific parts using the structure of the model. In figure 4.4 we present the graphs for the three stocks. As a general results information shares are highest at short durations. For INTC the dominance of Island is obvious. For CMGI MM2 dominates and for AAPL Instinet leads in terms of information shares. Another general result is that for CMGI the information share of a tick is much higher than that of the other two stocks.

#### 4.5.3 Calendar time aggregated Price Discovery

In the previous part we discussed some results for the information content of quote innovations. When these measures are aggregated over time, these provide useful information on the total efficiency of dealer as well as stocks and allow us to determine the total price contribution of dealers. Another benefit of time aggregation is that the measure we find can be compared to traditional measures, like Hasbrouck (1995) information shares. In this part we aggregate to one-minute and 5-minute intervals, as these intervals are most commonly used in other research.

We again start by discussing dealer efficiency. In table 4.5 we report these measures in 60 second intervals and 300 second intervals. Depending on the sign of  $\alpha$ , these measures are higher or lower than one. As mentioned when the time period over which we aggregate grows, these measures should converge to one. This convergence is clearly seen for all of the stocks in the sample. Many stocks have achieved this convergence after five-minutes, like INTC, but for the less liquid stocks there are still some inefficiencies (see e.g. QWST). In figure 4.5 we show the distributions of  $\beta(t)$  for the three stocks that we evaluate at 60 and 300 seconds. The results from these graphs again prove the heterogeneity among dealers. It also displays the skewed distribution of this measure.

The same is done for the measure of price discovery in table 4.6. A general result that can be seen is that price discovery increases when the sampling interval gets larger. At both sampling intervals we see that Island has very high measures for price discovery compared to the other dealers, especially for the more liquid stocks in the sample. These measures can again be shown in distribution plots in figure 4.6. We show these distributions for three stocks



Table 4.5: Average values for dealer efficiency ( $\beta(t)$ ) per dealer quote

Stock	$ISLD_b$	$ISLD_a$	$INCA_b$	$INCA_a$	$MM1_b$	$MM1_a$	$MM2_b$	$MM2_a$	$MM3_b$	$MM3_a$
60 Second intervals										
AAPL	0.89	0.94	1.13	1.06	1.34	1.22	1.11	0.84	0.91	0.98
AMAT	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
AMGN	0.81	0.92	1.02	1.02	0.83	0.89	0.80	0.80	1.02	0.98
AMZN	1.00	1.02	1.14	1.21	0.95	0.90	0.85	0.22	0.98	0.99
ATHM	1.00	1.00	1.00	1.00	1.01	1.01	1.00	1.00	1.00	1.00
CMGI	0.98	0.96	1.02	1.03	0.97	0.99	1.12	1.10	0.93	0.90
COMS	1.00	1.00	1.08	1.05	0.92	0.94	1.00	0.99	1.00	1.10
CPWR	0.99	0.99	0.99	1.00	1.03	1.02	1.04	1.04	1.01	0.99
CSCO	0.98	0.98	1.02	1.02	1.02	1.00	1.03	1.03	0.87	0.87
DELL	0.98	0.98	1.01	1.01	1.19	1.23	1.00	1.03	0.90	0.88
INTC	1.00	0.98	0.99	1.00	1.03	1.05	0.93	0.95	0.93	0.93
MSFT	1.00	1.00	0.99	1.00	0.95	0.99	0.92	0.93	0.83	0.84
NOVL	1.17	0.99	1.25	1.02	1.26	0.62	1.04	1.15	1.02	0.99
NXTL	0.82	0.84	0.94	1.00	1.24	1.16	1.19	0.83	1.02	0.81
ORCL	0.99	0.99	1.01	1.00	1.07	0.94	0.99	1.00	0.99	0.99
PSFT	1.16	0.93	1.33	1.03	3.05	2.63	2.39	2.58	0.96	1.40
QWST	1.52	1.11	3.39	3.85	0.76	1.12	2.46	2.80	1.37	1.19
SBUX	1.02	0.99	1.21	1.04	1.33	1.23	0.91	0.80	0.95	0.97
SUNW	1.00	0.99	1.00	1.00	0.99	0.99	0.97	0.96	1.00	1.00
WCOM	1.00	1.00	1.01	1.00	1.02	1.01	0.99	0.99	1.00	1.00
300 Second intervals										
AAPL	0.97	0.99	1.03	1.02	1.08	1.05	1.03	0.96	0.98	0.99
AMAT	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AMGN	0.96	0.98	1.00	1.00	0.96	0.97	0.95	0.95	1.00	1.00
AMZN	1.00	1.00	1.03	1.05	0.99	0.98	0.97	0.81	1.00	1.00
ATHM	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
CMGI	0.99	0.99	1.01	1.01	0.99	1.00	1.03	1.02	0.98	0.98
COMS	1.00	1.00	1.02	1.01	0.98	0.99	1.00	1.00	1.00	1.02
CPWR	1.00	1.00	1.00	1.00	1.01	1.00	1.01	1.01	1.00	1.00
CSCO	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.01	0.97	0.97
DELL	1.00	1.00	1.00	1.00	1.04	1.05	1.00	1.01	0.98	0.98
INTC	1.00	1.00	1.00	1.00	1.01	1.01	0.99	0.99	0.99	0.99
MSFT	1.00	1.00	1.00	1.00	0.99	1.00	0.98	0.99	0.97	0.97
NOVL	1.04	1.00	1.06	1.01	1.07	0.90	1.01	1.04	1.01	1.00
NXTL	0.96	0.96	0.98	1.00	1.06	1.04	1.05	0.96	1.01	0.95
ORCL	1.00	1.00	1.00	1.00	1.02	0.99	1.00	1.00	1.00	1.00
PSFT	1.04	0.98	1.08	1.01	1.47	1.38	1.32	1.36	0.99	1.09
QWST	1.12	1.03	1.54	1.65	0.95	1.03	1.33	1.41	1.08	1.04
SBUX	1.01	1.00	1.06	1.01	1.10	1.07	0.97	0.94	0.99	0.99
SUNW	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00
WCOM	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

*Note:* This table reports the averages for dealer efficiency ( $\beta(t)$ ) per dealer quote. This measure is defined as the covariance between quote innovations and the change in the efficient price, divided by the variance of the efficient price change. Panel A shows this measure aggregated up to 60 seconds, panel B up to 300 seconds. All measures over time converge to 1.

that we have selected and only show the asks. Clearly when the sampling interval gets larger the distributions converge to a point estimate. In contrast to information shares, our measure benefits from time aggregation. The longer the sampling interval the more precisely we can assign information shares uniquely. In the case of INTC and CMGI Island clearly dominates the price discovery process. For AAPL after five minutes there is still no convergence to fixed

Table 4.6: Average values for price discovery ( $\gamma(t)$ ) per dealer quote

Stock	ISLD <sub>b</sub>	ISLD <sub>a</sub>	INCA <sub>b</sub>	INCA <sub>a</sub>	MM1 <sub>b</sub>	MM1 <sub>a</sub>	MM2 <sub>b</sub>	MM2 <sub>a</sub>	MM3 <sub>b</sub>	MM3 <sub>a</sub>
60 Second intervals										
AAPL	0.03	0.02	0.02	0.02	0.02	0.02	0.04	0.03	0.01	0.02
AMAT	0.14	0.22	0.04	0.04	0.05	0.05	0.03	0.03	0.02	0.02
AMGN	0.06	0.06	0.03	0.02	0.06	0.06	0.03	0.02	0.05	0.04
AMZN	0.30	0.38	0.02	0.03	0.01	0.01	0.07	0.04	0.01	0.01
ATHM	0.13	0.13	0.04	0.04	0.06	0.05	0.04	0.04	0.02	0.03
CMGI	0.24	0.21	0.04	0.04	0.04	0.04	0.07	0.07	0.02	0.02
COMS	0.10	0.08	0.03	0.03	0.02	0.02	0.03	0.02	0.03	0.03
CPWR	0.03	0.03	0.03	0.03	0.07	0.06	0.04	0.04	0.02	0.02
CSCO	0.30	0.32	0.02	0.02	0.02	0.01	0.01	0.01	0.02	0.01
DELL	0.34	0.41	0.01	0.02	0.01	0.01	0.01	0.00	0.01	0.01
INTC	0.12	0.18	0.03	0.03	0.02	0.03	0.05	0.04	0.01	0.02
MSFT	0.16	0.14	0.03	0.03	0.02	0.02	0.03	0.02	0.02	0.02
NOVL	0.01	0.01	0.01	0.01	0.01	0.00	0.02	0.02	0.01	0.01
NXTL	0.01	0.00	0.03	0.02	0.03	0.02	0.02	0.01	0.04	0.02
ORCL	0.12	0.12	0.03	0.03	0.03	0.03	0.06	0.05	0.03	0.05
PSFT	0.02	0.03	0.02	0.02	0.08	0.09	0.00	0.00	0.02	0.01
QWST	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.05	0.05
SBUX	0.00	0.00	0.00	0.00	0.32	0.30	0.00	0.00	0.01	0.01
SUNW	0.11	0.13	0.04	0.04	0.03	0.02	0.04	0.04	0.03	0.02
WCOM	0.05	0.07	0.03	0.04	0.02	0.02	0.04	0.03	0.01	0.02
300 Second intervals										
AAPL	0.06	0.05	0.03	0.04	0.04	0.04	0.09	0.07	0.04	0.04
AMAT	0.19	0.30	0.05	0.05	0.07	0.07	0.04	0.05	0.02	0.03
AMGN	0.12	0.10	0.04	0.04	0.10	0.10	0.06	0.04	0.07	0.06
AMZN	0.32	0.41	0.03	0.03	0.02	0.02	0.08	0.05	0.01	0.01
ATHM	0.20	0.19	0.05	0.06	0.08	0.08	0.06	0.06	0.03	0.04
CMGI	0.29	0.25	0.04	0.04	0.05	0.05	0.08	0.08	0.03	0.03
COMS	0.17	0.15	0.06	0.04	0.04	0.04	0.05	0.04	0.06	0.05
CPWR	0.05	0.05	0.05	0.06	0.13	0.11	0.06	0.06	0.04	0.04
CSCO	0.37	0.40	0.03	0.03	0.02	0.02	0.01	0.01	0.03	0.02
DELL	0.39	0.48	0.01	0.02	0.01	0.01	0.01	0.00	0.02	0.01
INTC	0.19	0.28	0.05	0.05	0.04	0.04	0.08	0.07	0.02	0.02
MSFT	0.26	0.24	0.05	0.05	0.03	0.03	0.05	0.04	0.04	0.03
NOVL	0.02	0.02	0.02	0.02	0.02	0.01	0.04	0.04	0.03	0.04
NXTL	0.02	0.01	0.06	0.05	0.05	0.05	0.04	0.04	0.08	0.06
ORCL	0.18	0.18	0.05	0.05	0.05	0.04	0.09	0.08	0.05	0.07
PSFT	0.06	0.08	0.04	0.04	0.12	0.14	0.01	0.00	0.04	0.03
QWST	0.03	0.04	0.02	0.02	0.06	0.05	0.03	0.02	0.10	0.11
SBUX	0.00	0.00	0.01	0.01	0.41	0.38	0.01	0.01	0.01	0.01
SUNW	0.17	0.21	0.07	0.07	0.06	0.04	0.06	0.06	0.04	0.04
WCOM	0.11	0.14	0.06	0.07	0.04	0.04	0.07	0.07	0.03	0.03

Note: This table reports the averages for price discovery ( $\gamma(t)$ ) per dealer quote. This measure is obtained from the regression of the efficient price change on quote innovations. Panel A shows this measure aggregated up to 60 seconds, panel B up to 300 seconds.

Table 4.7: Averages for Information shares per dealer

Stock	60 seconds					300 seconds				
	Island	Instinet	MM1	MM2	MM3	Island	Instinet	MM1	MM2	MM3
AAPL	0.04	0.03	0.03	0.07	0.03	0.11	0.07	0.07	0.16	0.08
AMAT	0.35	0.07	0.10	0.06	0.04	0.50	0.10	0.14	0.09	0.06
AMGN	0.13	0.04	0.11	0.06	0.07	0.22	0.08	0.20	0.11	0.12
AMZN	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00
ATHM	0.26	0.07	0.11	0.08	0.05	0.39	0.11	0.16	0.12	0.07
CMGI	0.46	0.07	0.09	0.13	0.05	0.55	0.09	0.11	0.15	0.06
COMS	0.17	0.05	0.05	0.05	0.06	0.32	0.10	0.09	0.09	0.11
CPWR	0.05	0.06	0.13	0.07	0.04	0.10	0.10	0.24	0.12	0.07
CSCO	0.61	0.04	0.03	0.02	0.05	0.77	0.05	0.03	0.02	0.06
DELL	0.74	0.03	0.01	0.01	0.03	0.87	0.03	0.01	0.01	0.03
INTC	0.29	0.07	0.05	0.09	0.03	0.46	0.11	0.07	0.15	0.05
MSFT	0.28	0.05	0.03	0.05	0.04	0.49	0.09	0.05	0.09	0.08
NOVL	0.01	0.02	0.01	0.04	0.03	0.03	0.04	0.03	0.09	0.06
NXTL	0.01	0.04	0.04	0.03	0.06	0.03	0.11	0.09	0.08	0.14
ORCL	0.23	0.07	0.05	0.11	0.08	0.36	0.10	0.08	0.17	0.12
PSFT	0.00	0.00	0.98	0.00	0.00	0.00	0.00	1.00	0.00	0.00
QWST	0.03	0.03	0.04	0.03	0.09	0.07	0.07	0.09	0.06	0.21
SBUX	0.00	0.01	0.74	0.01	0.01	0.01	0.01	0.88	0.01	0.01
SUNW	0.23	0.08	0.06	0.07	0.05	0.39	0.14	0.09	0.12	0.08
WCOM	0.12	0.07	0.04	0.07	0.03	0.25	0.14	0.08	0.14	0.06

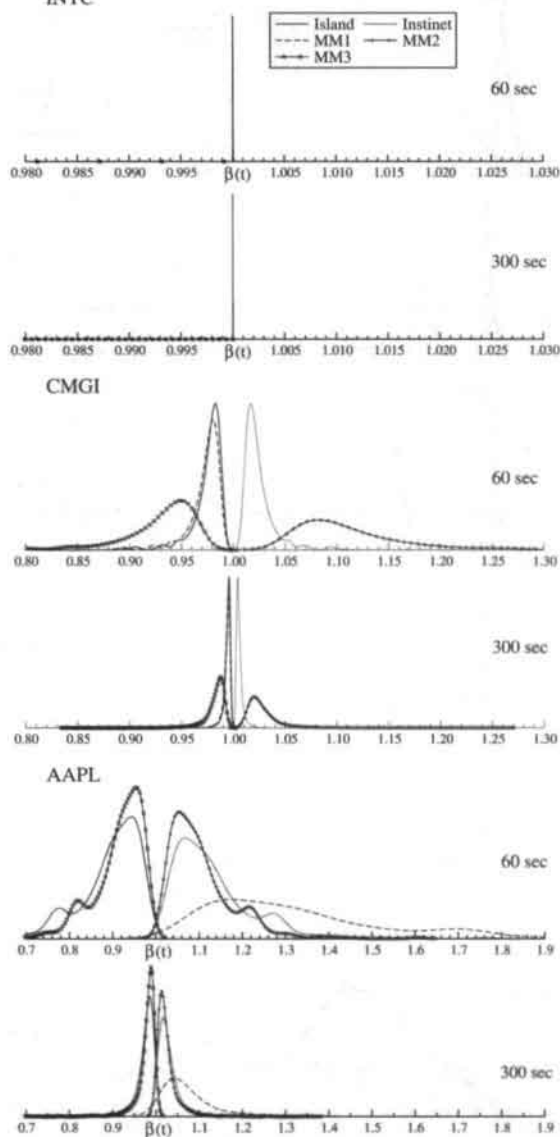
*Note:* This table reports the averages for information shares per dealer. This measure is obtained as the inner product of  $\lambda(t)$  and  $\theta(t)$ . The matrix inversion lemma is used to assign the information shares to separate dealers. This is possible as we impose structure on the idiosyncratic dealer noise. We reports the measure aggregated up to 60 seconds and 300 seconds.

points.

Finally we consider the  $R^2(t)$  of the regression when we aggregate over time. By applying the inversion lemma twice, this  $R^2(t)$  can be decomposed and information shares can be assigned uniquely. Results for information shares per dealer are shown in table 4.7. Again comparing the 60 and 300 second interval we see that information shares increase. Interesting is the total dominance of specific market makers for the stocks PSFT and SBUX. Another interesting result is that we never find any dominance Instinet. This is in contrast to the findings of Huang (2002). In figure 4.7 we show these  $R^2(t)$  per dealer for the three stocks. These show the same pattern as the graphs for  $\gamma(t)$ . This is because the measure for  $\beta(t)$  converge to one and the  $R^2(t)$  is the sum of the elements in  $\gamma(t)$ .

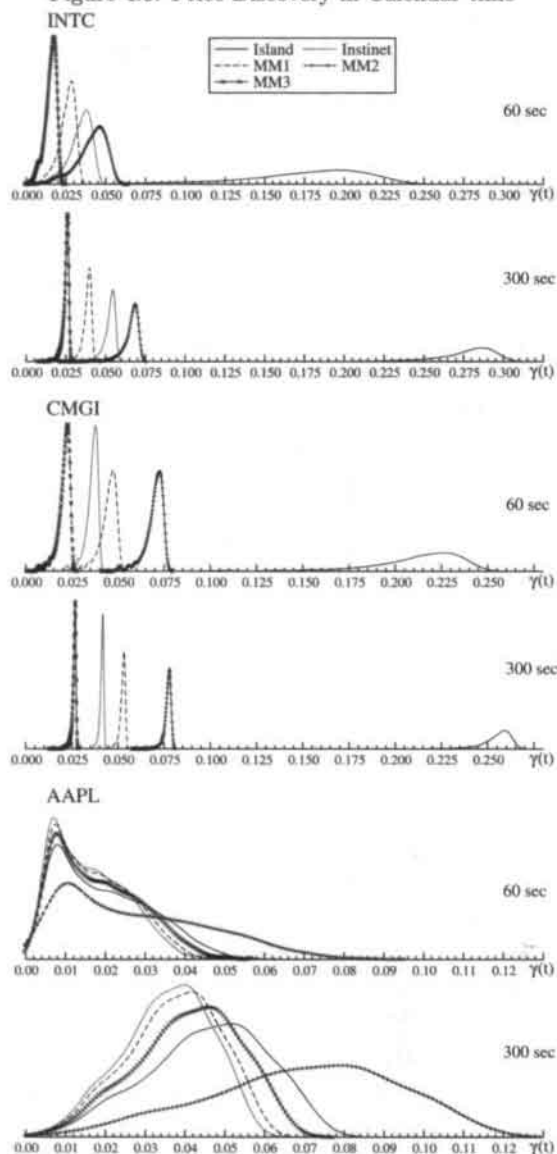
In figure 4.8 we show the total sum of the  $R^2(t)$  for a one-, two- and five-minute interval. With these graphs we can discuss the efficiency of the stock as a whole. When we move to longer sampling intervals more information is incorporated in the efficient price of the stock. In the case of INTC and CMGI, this information is almost fully incorporated after five minutes. In the case of AAPL we see that after five minutes there is still a lot of specific

Figure 4.5: Dealer efficiency in Calendar time  
INTC



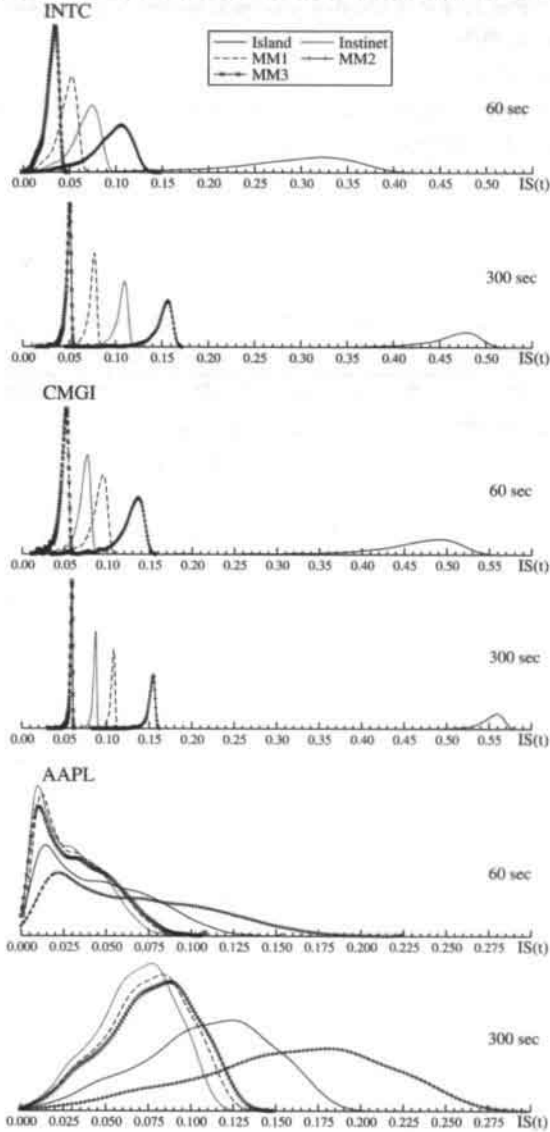
*Note:* These graphs show the distributions of the calendar time measures for dealer efficiency. These distributions are shown for the bids of INTC, CMGI and AAPL. Island is not shown as its value is always 1. The aggregates are shown at 60 and 300 second intervals.

Figure 4.6: Price Discovery in Calendar time



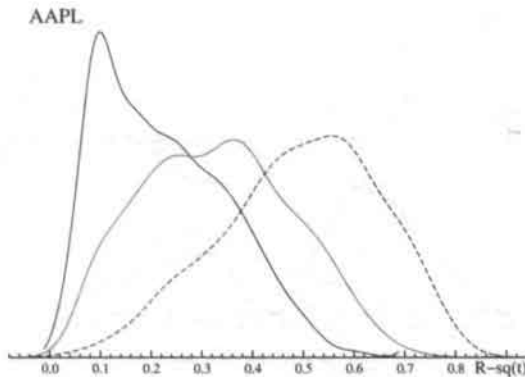
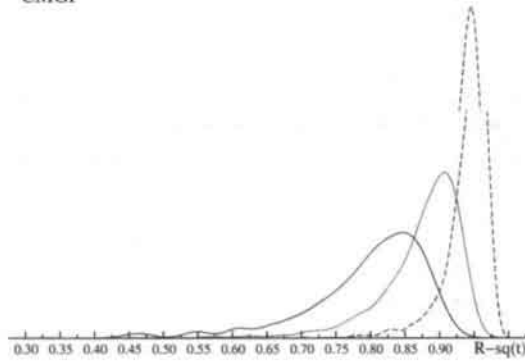
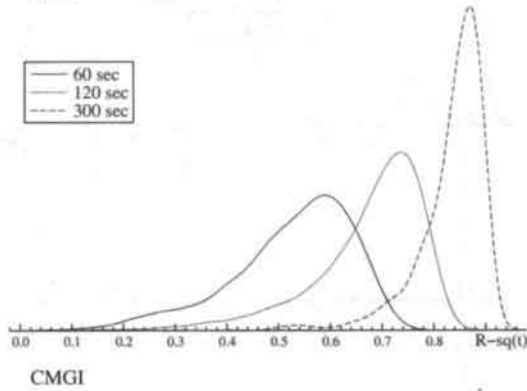
*Note:* These graphs show the distributions of the calendar time measures for price discovery. These distributions are shown for the asks of INTC, CMGI and AAPL. The aggregates are shown at 60 and 300 second intervals.

Figure 4.7: Information Shares in Calendar time per dealer



*Note:* These graphs show the distributions of the calendar time measures for Information Shares. These distributions are shown for INTC, CMGI and AAPL. The aggregates are shown at 60 and 300 second intervals. This measure is obtained by taking the inner product of the measure for dealer efficiency and price discovery. Using the specific structure of the model and by applying the matrix inversion lemma, the inner product can be decomposed to dealers.

Figure 4.8: Calendar time aggregated Information  
INTC



*Note:* These graphs show the distributions of the total information incorporated in dealer quotes over a specific time interval. These distributions are shown for INTC, CMGI and AAPL. The aggregates are shown at 60, 120 and 300 second intervals.

dealer noise.

Implications of this also run into a different field of research. When measuring realized volatility of Andersen, Bollerslev, Diebold, and Labys (2001) one usually samples at 5-minute intervals. For the most liquid stocks this seems to be a good sampling interval as most information is incorporated in the stock. However for the less liquid stock microstructure effects can still dominate the movement of the stock price and thus the assumed convergence to the total integrated volatility does not hold.

## 4.6 Conclusion

This chapter introduced a model for dealer quoting behavior in tick time. Quote innovations are modelled as they arrive and the model can be estimated using a Kalman Filter. From the model we derived measures for dealer efficiency and price discovery in tick time. Consequently these measures were aggregated to calendar time equivalents. We showed that when these measures are aggregated over a sufficiently long period of time, these will converge to single point estimates for price discovery.

In our empirical results we find that more volatility is generated at shorter durations. We also find that dealer quotes tend to be more informative when durations are short. Using the measures for price discovery we can identify the information share of the dealer examined.

Finally the results for the calendar time aggregation, indicate that for some stocks, the microstructure noise in the price process is still relatively large. This implies that at a sampling interval of five minutes quotes are still noisy. Hence measures like realized volatility, for some stocks might still be noise approximations of the real volatility.



## 4.A Appendix: Derivation of the Moment Conditions

This appendix discusses the identification problems that can arise. The assumption that there will be only one quote update leads to the restriction that  $J'_\ell = 1$ . To make this point consider no variation in durations. Hence the duration functions can be left out for the moment. Constants are left out for the sake of notational clarity. Consider model (4.5) in section 4.2.

$$\begin{aligned} J'_\ell q_\ell &= J'_\ell m_\ell + J'_\ell u_\ell, \\ m_\ell &= m_{\ell-1} + \sigma r_\ell, \\ J'_\ell u_\ell &= J'_\ell \alpha \sigma r_\ell + J'_\ell e_\ell, \end{aligned} \tag{4.25}$$

Given the structure of this model we can derive the variance and the first order auto covariance,

$$\begin{aligned} E[\Delta q_\ell^2] &= \gamma_0(t_\ell) = \\ &= (1 + 2J'_\ell \alpha + J'_\ell \alpha \alpha' J_\ell + J'_{\ell-1} \alpha \alpha' J_{\ell-1}) \sigma^2 \\ &\quad + J'_\ell \Omega J_\ell + J'_{\ell-1} \Omega J_{\ell-1}, \end{aligned} \tag{4.26}$$

$$\begin{aligned} E[\Delta q_\ell, \Delta q_{\ell-1}] &= \gamma_1(t_{\ell-1}) = \\ &= -\sigma^2 (J'_{\ell-1} \alpha + J'_{\ell-1} \alpha \alpha' J_{\ell-1}) \\ &\quad - J'_{\ell-1} \Omega J_{\ell-1}. \end{aligned} \tag{4.27}$$

Note that the variance is driven only by the information in the last observation and that the first order auto covariance is driven by the lagged information. To show that  $\alpha$  cannot be identified uniquely, define

$$\dot{\alpha} \equiv \alpha + x\ell \tag{4.28}$$

The moment conditions can be rewritten as

$$\begin{aligned} \gamma_0(t_\ell) &= \sigma^2 (1 + 2J'_\ell \dot{\alpha} + J'_\ell \dot{\alpha} \dot{\alpha}' J_\ell + J'_{\ell-1} \dot{\alpha} \dot{\alpha}' J_{\ell-1}) \\ &\quad + \sigma^2 (2x + x \dot{\alpha}' J_\ell + x J'_\ell \dot{\alpha} + x \dot{\alpha}' J_{\ell-1} + x J'_{\ell-1} \dot{\alpha} + 2x^2) \\ &\quad + J'_\ell \dot{\Omega} J_\ell + J'_{\ell-1} \dot{\Omega} J_{\ell-1} \end{aligned} \tag{4.29}$$

$$\begin{aligned} \gamma_1(t_{\ell-1}) &= -\sigma^2 (J'_{\ell-1} \dot{\alpha} + J'_{\ell-1} \dot{\alpha} \dot{\alpha}' J_{\ell-1}) \\ &\quad - \sigma^2 (x + x J'_{\ell-1} \dot{\alpha} + x \dot{\alpha}' J_{\ell-1} + x^2) \\ &\quad - J'_{\ell-1} \dot{\Omega} J_{\ell-1}. \end{aligned} \tag{4.30}$$

Which can be solved when

$$\dot{\Omega} = \Omega + \sigma^2 x u' (\mu' + \dot{\alpha} \mu' + \mu \dot{\alpha}' + x \mu'). \quad (4.31)$$

Hence the model stated cannot identify the parameters in the  $\alpha$  vector uniquely. When durations dependence is assumed in the model, the model reads

$$\begin{aligned} J'_\ell q_\ell &= J'_\ell \mu m_\ell + J'_\ell u_\ell, \\ m_\ell &= m_{\ell-1} + \sigma_\ell r_\ell, \\ J'_\ell u_\ell &= J'_\ell \alpha \tau_\ell^{\delta_2} \sigma r_\ell + J'_\ell e_\ell, \end{aligned} \quad (4.32)$$

where

$$\sigma_\ell = \sigma \tau_\ell^{\delta_1}. \quad (4.33)$$

Again moments can be derived for this model:

$$\begin{aligned} \gamma_0(t_\ell) &= \sigma^2 (\tau_\ell^{2\delta_1} + 2\tau_\ell^{\delta_1+\delta_2} J'_\ell \alpha + \tau_\ell^{2\delta_2} J'_\ell \alpha \alpha' J_\ell + \tau_{\ell-1}^{2\delta_2} J'_{\ell-1} \alpha \alpha' J_{\ell-1}) \\ &\quad + J'_\ell \Omega J_\ell + J'_{\ell-1} \Omega J_{\ell-1} \end{aligned} \quad (4.34)$$

$$\begin{aligned} \gamma_1(t_{\ell-1}) &= -\sigma^2 (\tau_{\ell-1}^{\delta_1+\delta_2} J'_{\ell-1} \alpha + \tau_{\ell-1}^{2\delta_2} J'_{\ell-1} \alpha \alpha' J_{\ell-1}) \\ &\quad - J'_{\ell-1} \Omega J_{\ell-1}. \end{aligned} \quad (4.35)$$

The vector  $\alpha$  can now be identified uniquely iff durations change over time. When these durations are fixed a linear substitute can be found in the way shown before and the vector  $\alpha$  cannot be identified.



## Chapter 5

# Realized Variance in the Presence of Non-IID Microstructure Noise<sup>1</sup>

### 5.1 Introduction

Probably one of the most rapidly developing fields in finance is the area of realized variance ( $RV$ ). Introduced by Andersen, Bollerslev, Diebold, and Labys (2001),  $RV$ , defined as the sum of squared intraday returns, converges to the integrated variance ( $IV$ ) when the sampling frequency goes to infinity (see Barndorff-Nielsen and Shephard (2002) and Meddahi (2002)). This consistency relies on the assumption that a market operates without frictions and that there are no microstructure effects present. But when sampling frequencies are very high, prices are contaminated by microstructure noise and  $RV$  can be biased and inconsistent for  $IV$  (see e.g. Andreou and Ghysels (2002) and Oomen (2002)). This issue is circumvented by sampling at lower frequencies, where the effects of microstructure noise are negligible. Although sampling at lower frequencies reduces the bias in  $RV$  introduced by microstructure noise, it increases the measurement error of  $RV$ . Finding the optimal sampling frequency results in finding the best trade-off between bias and variance and is the approach followed by e.g. Bandi and Russell (2003). Zhang, Mykland, and Ait-Sahalia (2003) propose an alternative measure for  $RV$  based on sub-sampling techniques which reduces the loss of information inherent in data aggregation and at the same time provides a consistent estimator for  $IV$ . However, all these approaches rely on data aggregation to mitigate impact of the microstructure noise.

In this chapter we filter the microstructure noise from the “efficient” price process, while sampling at the highest frequency. We follow several approaches to this decomposition. First we deal with the microstructure noise in transaction time, or so called tick time. For this tick

---

<sup>1</sup>This Chapter is based on B. Frijns and T. Lehnert, ‘Realized Variance in the Presence of non-IID Microstructure Noise: A Structural Approach’, LIFE Working Paper WP04-008.

time data, we propose a structural time series decomposition of the observed price into an efficient price and microstructure noise. The traditional assumption is that the microstructure noise is i.i.d. (e.g. Bandi and Russell (2003), Hansen and Lunde (2004a) and Zhang, Mykland, and Ait-Sahalia (2003)). However, recent research has shown that this is not the case (see e.g. Hansen and Lunde (2004b)). We therefore allow the microstructure noise to correlate with the efficient price and to depend on the time of the day. Second, we follow a recently introduced model-free correction (Hansen and Lunde (2004b)). We apply this correction to data sampled at every second (sampling in calendar time) and data sampled at every transaction (sampling in tick time).

Applying the structural time series model, we find a realized variance that is substantially lower than the realized variance observed at lower frequencies. Further applying this structural model at several different sampling frequencies reveals that  $RV$  increases when the sampling frequency decreases. The structural model provides some additional insights. We find that the microstructure noise is highly correlated with the innovation in the efficient price. We also confirm that the microstructure noise is not distributed identically over the day. Microstructure noise is highest at the start of the trading day, is lowest during the afternoon and increases again towards the close of the market. These findings are in line with Madhavan, Richardson, and Roomans (1997).

As we sample in transaction time, we observe the durations between transactions. This allows us to test for the impact of time between transactions on the innovation in the efficient price. Our results suggest that for the evolution of the efficient price, calendar time is not the appropriate time scale. We additionally test for the impact of time between trades on the correlation between the microstructure noise and the innovation in the efficient price. At the highest sampling frequency these durations have a slightly positive impact, but this impact becomes negative at lower frequencies.

To further address the issue of  $RV$  we apply a recently introduced model-free correction (see Hansen and Lunde (2004b)). This approach corrects for the autocorrelations in observed returns, induced by the microstructure noise. The best<sup>2</sup> measure of  $RV$  is found at the highest sampling frequency, however the lag length chosen to correct over is rather arbitrary. We apply this correction in both time scales, when sampling is done in calendar time and when sampling is done in tick time.

Both corrections provide new insights in the properties of  $RV$ . First, when we apply the correction in calendar time on a one second sampling frequency, we find a higher value for  $RV$  (compared to lower frequency  $RV^3$ ) when the correction is done using only few lags.

<sup>2</sup>Best means that it minimizes the mean squared error from the true volatility.

<sup>3</sup>For the low frequency  $RV$  we use the realized variance at a 30 minute sampling frequency. At this frequency the effects of microstructure noise should be negligible. Hansen and Lunde (2004b) use the same benchmark for  $RV$

This measure decreases when the lag length increases, where the lag length needed to reach low frequency *RV* depends on the liquidity of the asset (about 5 seconds for the most liquid stocks and up to 5 minutes for the least liquid stocks). The correction done in tick time reveals a different pattern. Correcting for the first lag removes all or most of the negative autocorrelation in observed returns and finds a value for *RV* close to the value found with the structural approach. Correcting for longer lags, increases *RV*. However, up to 30 lags are needed to reach a level of *RV* equal to the low frequency *RV*.

These findings suggest that there are two factors that affect the efficient price process. The first one is the known microstructure noise, which generates negative autocorrelations and has only a short term effect (when sampling in tick time). The second is a long persistent positive autocorrelation, in tick time, of up to 30 lags. Hasbrouck and Ho (1987) find some persistence in transaction returns, but only up to five lags. Microstructure theory can explain some persistence in transaction returns, due to the presence of informed traders (see Glosten and Harris (1988)), but cannot explain this long persistence in observed transaction returns. We therefore question whether this persistency is a source of microstructure noise or reveal a momentum effect present in high frequency data.

The remainder of this chapter is structured as follows. In the next section we discuss the concept of realized volatility. We discuss realized volatility with sampling in different time scales. Subsequently, section 5.3 addresses microstructure noise and the effects it has on observed prices. In this section we address a structural approach to estimating realized variance and we consider a model-free correction on data sampled in tick time and data sampled in calendar time. Section 5.4 discusses the data used in this study. In section 5.5 we present the results for realized variance applying different ways of sampling and different corrections. Finally, section 5.6 concludes.

## 5.2 Realized Variance

In this section we introduce the concept of Realized Variance proposed by Andersen, Bollerslev, Diebold, and Labys (2001). We start with the case where no microstructure noise is present and assume the price process evolves in calendar time. Consecutively, we show that Realized Variance is the same whether sampling is done in calendar time or tick time.

Realized Variance is the discrete time approximation to the Integrated Variance (*IV*). This approximation holds under the assumption that the continuous price process is a semimartingale. We first define *IV* and then *RV*.

Let  $p_t$  be the log price of an asset at time  $t$  and assume that the return process follows a stochastic process of the form

$$dp_t = \mu_t dt + \sigma_t dW_t, \quad (5.1)$$

where  $\mu_t$  is a drift term of bounded variation,  $\sigma_t$  the spot volatility process and  $W_t$  is a standard Brownian process. The spot volatility process is assumed cadlag, bounded away from zero and independent of  $W_t$ . If we assume that  $t$  is measured in days, then the return from  $t-1$  to  $t$  is given by the integral

$$p_t - p_{t-1} = \int_{t-1}^t \mu_u du + \int_{t-1}^t \sigma_u dW_u. \quad (5.2)$$

The Integrated Variance of this day is defined by

$$IV_{[t-1,t]} = \int_{t-1}^t \sigma_u^2 du, \quad (5.3)$$

which also defines the continuous quadratic variation (see e.g. Revuz and Yor (1999)).

Realized Variance is the discrete time approximation to the Integrated Variance and is defined as the sum of squared intraday returns. In calendar time these returns are sampled at equidistant intervals. On day  $t-1$  we define the  $i^{th}$  intraday calendar time return as

$$r_{C,t-1+ih}^{(h)} = p_{t-1+ih} - p_{t-1+(i-1)h}, \quad (5.4)$$

where  $i = 1, \dots, h^{-1}$  and  $h^{-1}$  is a positive integer referring to the intraday sampling frequency. The superscript  $(h)$  is used to indicate the frequency sampled at. The subscript  $C$  refers to the fact that sampling is done in calendar time.

The Realized Variance in calendar time sampling on day  $t$  is defined by

$$RV_{C,t}^{(h)} \equiv \sum_{i=1}^{h^{-1}} \left\{ r_{C,t-1+ih}^{(h)} \right\}^2. \quad (5.5)$$

By the theory of quadratic variation it holds that  $RV$  is a consistent measure of  $IV$  (see e.g. Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shephard (2002)), i.e. when  $h \rightarrow 0$ ,

$$\text{plim} \sum_{i=1}^{h^{-1}} \left\{ r_{C,t-1+ih}^{(h)} \right\}^2 \rightarrow \int_{t-1}^t \sigma_u^2 du. \quad (5.6)$$

This convergence, however, only holds when returns are calculated in a correct manner. As sampling is done in calendar time, not every point sampled at will have a corresponding observation. The way these "missing observations" are filled in determines whether  $RV$  converges to  $IV$ . Hansen and Lunde (2004b), for example, propose two different methods to fill in these observations. The first method is referred to as the previous tick method (introduced by Wasserfallen and Zimmermann (1985)). This method records the previous price at every sampling point until a new price appears. For very high frequency returns this

means that many returns are equal to zero, with some nonzero returns when prices change. Sampling in this manner,  $RV$  provides a consistent estimate for  $IV$ . A second method often employed is linear interpolation. In this case, if no new price is observed at a sampling point, the observation is interpolated from the previous and the next price. Hansen and Lunde (2004b) show that in this case  $RV$  converges to zero in probability and thus does not provide a consistent measure for  $IV$ .

Additional to sampling in calendar time, one can also sample in tick time. When sampling in tick time we basically ignore the in-between zero returns and only observe the price when a transaction has taken place. Returns in transaction time are defined as follows. On day  $t - 1$  a total of  $k(t - 1)^{-1}$  transactions are observed, where  $j = 1, \dots, k(t - 1)^{-1}$ . The  $j^{th}$  intraday return is defined as

$$r_{T,t-1+jk(t)} = p_{t-1+jk(t)} - p_{t-1+(j-1)k(t)}. \quad (5.7)$$

The subscript  $T$  in this case refers to sampling in tick time. Notice that these returns are spaced equidistantly in tick time, but are spaced irregularly in calendar time. The Realized Variance in tick time is given by

$$RV_{T,t} \equiv \sum_{j=1}^{k(t)^{-1}} \{r_{t-1+jk(t)}\}^2. \quad (5.8)$$

An interesting feature of  $RV_{T,t}$  is that at the highest sampling frequency, it is the same as  $RV_{C,t}$  at the highest sampling frequency, when the previous tick rule is applied for the calendar time observations. If, for example, we record prices at every second using a previous tick rule, zero returns are recorded if prices do not change. When summing the squared returns, only the non-zero returns determine  $RV$ . These non-zero returns are exactly the same as the returns that we sample in tick time, by sampling at every tick. This entails that both methods provide the same consistent measure for  $IV$ .

### 5.3 Microstructure Noise

In the previous section we addressed  $RV$  when no microstructure noise is present and have shown that  $RV$  is the same whether it is determined in calendar time or in tick time. In this section we continue in discrete time to discuss microstructure noise and its impact on  $RV$ . We further address the impact of microstructure noise on the time series properties of observed returns in both time scales. Consecutively, we propose a structural approach to filter out the microstructure noise. Finally, we address a new method to correct for the microstructure noise.



### 5.3.1 The Effects of Microstructure Noise on Realized Variance

Although  $RV$  converges to  $IV$  when the sampling frequency goes to infinity, the presence of microstructure noise at these frequencies seriously contaminates the observed price process. This causes  $RV$  to be a biased measure for  $IV$ . Without making any reference to a particular time scale the observed price at any given time  $0 < d < 1$  on day  $t$  can be decomposed as

$$p_{t+d} = p_{t+d}^* + u_{t+d}, \quad (5.9)$$

where  $p_{t+d}^*$  refers to the latent *true* price and  $u_{t+d}$  refers to the microstructure noise, which is also latent. The true price is often referred to as the efficient price and as such is often assumed to follow a random walk. The microstructure noise, on the other hand is considered transitory and is often assumed to be a covariance stationary process. Moreover, the microstructure noise is often assumed to be i.i.d. (see e.g. Bandi and Russell (2004), Hansen and Lunde (2004a) and Zhang, Mykland, and Aït-Sahalia (2003)).

One of the most trivial ways to circumvent the issue of microstructure noise is by aggregating data. Frictions, inherent to trading mechanisms have no permanent impact on prices. Therefore, if prices are observed at low frequencies the impact of microstructure noise on observed returns is negligible. This has been the first approach to circumvent the problems of dealing with microstructure noise (see e.g. Andersen, Bollerslev, Diebold, and Labys (2001) and Andersen (2004)), where sampling was done at five minute intervals. However, the sampling interval of five minutes is chosen rather arbitrarily and several other studies have addressed the issue of finding the correct sampling frequency. The correct frequency here entails finding the best trade-off between the bias introduced by the microstructure noise and the measurement error for  $RV$  introduced by data aggregation (see e.g. Bandi and Russell (2004)). The microstructure noise is minimized at the lowest possible frequency, whereas the measurement error of  $RV$  is minimized at the highest possible frequency.

An alternative approach is adopted by Zhang, Mykland, and Aït-Sahalia (2003). Their approach, based on sub-sampling techniques, leads to a lower measurement error in  $RV$  without introducing a bias due to microstructure noise. The idea of this approach is to sample at lower frequencies (e.g. five minutes), while moving the sampling window forward with small steps. This allows all the information to be incorporated in the estimate of volatility without suffering from the effects of microstructure noise.

These approaches all circumvent dealing with microstructure noise by relying on data aggregation in some way, without explicitly addressing the properties of the microstructure noise and its impact on observed returns. To address the impact for observed returns assume that sampling is done in calendar time and prices change at every observation. The observed

price can be written as

$$p_{t-1+ih} = p_{t-1+ih}^* + u_{t-1+ih}, \quad (5.10)$$

and returns in calendar time are given by

$$r_{C,t-1+ih}^{(h)} = (p_{t-1+ih}^* - p_{t-1+(i-1)h}^*) + (u_{t-1+ih} - u_{t-1+(i-1)h}), \quad (5.11)$$

which is a MA(1) process.

Assuming that the microstructure noise is i.i.d. and independent of the innovation in the efficient price, the variance of returns is given by

$$E[\{r_{C,t-1+ih}^{(h)}\}^2] = \sigma^{(h),2} + 2\omega^2, \quad (5.12)$$

where  $\sigma^{(h),2}$  is the variance of the true price and  $\omega^2$  is the variance of the microstructure noise.

If we further assume that the innovations in the efficient price are independent, then the microstructure noise introduces first order autocorrelation in observed returns,

$$E[\{r_{C,t-1+ih}^{(h)}\}\{r_{C,t-1+(i-1)h}^{(h)}\}] = -\omega^2. \quad (5.13)$$

All higher order autocorrelations are equal to zero. The last two equations (5.12) and (5.13) reveal that the first order autocorrelation in returns has a lower bound of  $-0.5$ .

In this example we have assumed that at each observation the price changes, and we expect to observe a MA(1) structure in returns if the sampling frequency is low enough to ensure this. But if calendar time sampling is done using the previous tick rule, then when no price change is observed the previous price is recorded. This causes the variance of returns in (5.12) to be either 0 or  $\sigma^{(h),2} + 2\omega^2$  and seriously affects the first order autocorrelation in (5.13). When zero returns are recorded the first order autocorrelation will be higher than that of a pure MA(1) structure (this issue was addressed by De Jong, Mahieu, and Schotman (1998)). As the observed price in this way of sampling changes after a longer time period, sampling using the previous tick rule will induce higher order negative autocorrelations in returns. These higher order autocorrelations will depend on the distribution of durations between observations.

On the other hand, if sampling is done in tick time, we do not observe these zero return observations. As we observe only prices when they change, sampling in this time scale will not induce higher order negative autocorrelations, assuming i.i.d.-ness of the microstructure noise. This time scale should reveal the MA(1) structure and therefore, in our opinion, is the more appropriate sampling time scale.

Further complications arise when the microstructure noise is not i.i.d. Firstly, the mi-

crostructure noise can be correlated with the efficient price process. Further, the size of the microstructure noise can depend on the time of the day (see e.g. Madhavan, Richardson, and Roomans (1997)). In the next subsection we introduce a model, for tick time data, that decomposes the microstructure noise from the efficient price process. We assume that the microstructure noise is independently distributed, but allow for some time of the day variation. We further allow the microstructure noise to correlate with the innovations in the efficient price.

### 5.3.2 A structural model for the price process

In this section we propose a structural model for the price process of a security that uses tick time data. We decompose the log price into two components: the efficient price and microstructure noise (see e.g. Hasbrouck (1993), Aït-Sahalia, Mykland, and Zhang (2003) and Zhang, Mykland, and Aït-Sahalia (2003) among others). We allow the microstructure noise to depend on the time of the day and allow for correlation between the microstructure noise and the innovation in the efficient price. In this model we consider all informative transactions (the next section elaborates on what we refer to as informative transactions) at the time they are reported to the trading system. We thus sample each transaction and do not aggregate the data to some equidistant sampling interval. For sake of clarity we omit the  $(a)$  superscript for the sampling frequency and assume that sampling is done at every tick. We will later reintroduce the superscript when needed.

Consider the price process observed in tick time. This price process is decomposed in a latent component referring to the efficient price and microstructure noise

$$p_{t+jk(t)} = p_{t+jk(t)}^* + u_{t+jk(t)}, \quad (5.14)$$

where the first component  $(p_{t+jk(t)}^*)$  refers to the true, or efficient price of an asset and the second component  $(u_{t+jk(t)})$  measures the microstructure noise in each transaction. The efficient price is assumed to follow a random walk. However, we let the model determine which time scale the price process evolves in. This is done by including a duration function in front of the innovation term of the random walk,

$$p_{t+jk(t)}^* = p_{t+(j-1)k(t)}^* + \tau_{t+jk(t)}^{\delta_{1,t}} \varepsilon_{t+jk(t)}, \quad (5.15)$$

where

$$\tau_{t+jk(t)} \equiv \frac{\ell_{t+jk(t)} - \ell_{t+(j-1)k(t)}}{\frac{1}{N} \sum_t k(t) \sum_j (\ell_{t+jk(t)} - \ell_{t+(j-1)k(t)})}$$

measures the deviation from the average duration over the sample, where  $N$  is the total number of days evaluated. This normalization only affects the innovation term in the efficient

price, which is now expressed in terms of ticks instead of seconds. The innovation term  $\varepsilon_{t+jk(t)}$  measures the innovation in the random walk at average durations and has a variance of  $\sigma_t^2$ . The parameter  $\delta_{1,t}$  measures the impact duration has on the innovation in the random walk. When  $\delta_{1,t} = \frac{1}{2}$  the variance of the random walk innovation grows proportional to the duration between observation. In this case the price process evolves in calendar time, where the variance of the innovation in the efficient price grows proportional to the time length between observation. This process would be the discrete time equivalent of the continuous price process in (5.1). On the other hand when  $\delta_{1,t} = 0$ , the time between observations does not influence the innovation in the random walk. In this case the variance of the random walk grows proportional to the number of transactions. This is the time scale that has been promoted more recently by Ané and Geman (2000).

We further want to specify the microstructure noise  $u_{t+jk(t)}$ . One factor that we consider is whether the innovation in the efficient price is correlated with the microstructure noise. Hansen and Lunde (2004b) stress the importance of the correlation that exists between microstructure noise and the innovation in the efficient price. Moreover, by including a parameter that measures the impact of time between transactions on these correlations we allow for richer dynamics than when sampling at fixed time intervals. We include a parameter that measures this correlation and we consider the effects of the opening and closing of the market. The amount of microstructure noise might differ substantially around the open and close of the market (see e.g. Madhavan, Richardson, and Roomans (1997)). Combining these two factors we specify the noise as

$$\begin{aligned} u_{t+jk(t)} = & \alpha_t \tau_{t+jk(t)}^{\delta_{2,t}} \varepsilon_{t+jk(t)} + D_{1,t+jk(t)} e_{1,t+jk(t)} \\ & + D_{2,t+jk(t)} e_{2,t+jk(t)} + D_{3,t+jk(t)} e_{3,t+jk(t)}, \end{aligned} \quad (5.16)$$

where  $\alpha_t$  measures the extent to which the innovation in the efficient price and the microstructure noise co-vary at average durations. The parameter  $\delta_{2,t}$  measures the impact that time has on this correlation. When  $\delta_{2,t} > 0$  the dependence increases with an increase in durations and vice versa. The three dummy variables  $D_{1,t+jk(t)}$ ,  $D_{2,t+jk(t)}$  and  $D_{3,t+jk(t)}$  are equal to zero or one, such that the three idiosyncratic noise terms ( $e_{1,t+jk(t)}$ ,  $e_{2,t+jk(t)}$ ,  $e_{3,t+jk(t)}$ ) have variances  $\omega_{1,t}^2, \omega_{2,t}^2, \omega_{3,t}^2$ , respectively and capture the microstructure noise at different times of the day. The dummy variable  $D_{1,t+jk(t)}$  equals 1 near the opening of the market (between 9.30 and 11.00) and zero otherwise,  $D_{2,t+jk(t)}$  equals 1 over the normal part of the trading day (11.00 - 14.30) and zero otherwise, and  $D_{3,t+jk(t)}$  equals 1 near the close (14.30 - 16.00) and zero otherwise.

The model can be put into a state space model and estimated by QML using a Kalman

Filter<sup>4</sup>

$$\begin{aligned}
p_{t+jk(t)} &= p_{t+jk(t)}^* + u_{t+jk(t)}, \\
p_{t+jk(t)}^* &= p_{t+(j-1)k(t)}^* + \tau_{t+jk(t)}^{\delta_{1,t}} \varepsilon_{t+jk(t)}, \\
u_{t+jk(t)} &= \alpha_t \tau_{t+jk(t)}^{\delta_{2,t}} \varepsilon_{t+jk(t)} + D_{1,t+jk(t)} e_{1,t+jk(t)} + \\
&\quad D_{2,t+jk(t)} e_{2,t+jk(t)} + D_{3,t+jk(t)} e_{3,t+jk(t)}.
\end{aligned} \tag{5.17}$$

As the latent variable in the state equation is a random walk the model is initialized using a diffuse prior. This entails that the initial prediction error variance is set at a very large number. For the prediction error variance to converge to normal levels we leave out the first 50 observations for the calculation of the likelihood function. As the innovation in the random walk depends on the time length between observations, the model does not converge to a steady state. This increases the computational efforts of this filter as the prediction error variance needs to be computed at every single recursion. All parameters in the model are identified, as long as  $\delta_2 \neq 0$  and  $\delta_1 + \delta_2 \neq 0$  and the variance of durations is not equal to zero (See the Appendix for a full derivation of the moment conditions).

In a first step, we estimate the model assuming constant parameters over all trading days. In this case, the subscript  $t$  can be omitted for the parameters. Every day, we re-initialize the system by increasing the prediction error variance and leaving out the first 50 observations on each day in the calculation of the likelihood function. This procedure allows us to analyze the properties of the microstructure noise effectively. As a second step, the model is re-estimated on a daily basis. In this case, the re-initialization procedure does not allow us to calibrate the model at lower than the highest frequency. In addition to the full calibration, this approach may capture more of the daily dynamics in  $RV$ .

Given the parameterization in the model above we can define the realized variance for the price process  $p_{t+jk(t)}$ ,

$$\begin{aligned}
RV_{T,t} &= \sum_{j=1}^{k(t)-1} (\Delta p_{t+jk(t)}^* + u_{t+jk(t)} - u_{t+(j-1)k(t)})^2 \\
&= \sum_{j=1}^{k(t)-1} \tau_{t+jk(t)}^{2\delta_{1,t}} \sigma_t^2 + 2\alpha_t^2 \sigma_t^2 \tau_{t+jk(t)}^{2\delta_{2,t}} + 2\omega_{\cdot,t} + \alpha_t^2 \sigma_t^2 \tau_{t+jk(t)}^{(\delta_{1,t}+\delta_{2,t})},
\end{aligned} \tag{5.18}$$

where  $\omega_{\cdot,t}$  is the variance of the idiosyncratic noise component at either opening, middle or

<sup>4</sup>For more information on state space models and Kalman Filter techniques, we refer to Harvey (1989) and Durbin and Koopman (2001).

end of the day. The first part of the previous equation,

$$RV_{T,p^*,t} = \sum_{j=1}^{k(t)-1} \tau_{t+jk(t)}^{2\delta_{1,t}} \sigma_t^2, \quad (5.19)$$

is an unbiased measure for the  $RV$  of the efficient price. The remaining part in (5.18) is the bias in  $RV$  introduced by the microstructure noise. In the empirical part we estimate the model at different sampling frequencies and compute  $RV$  at these frequencies.

### 5.3.3 A Model Free Correction

The previous subsection introduced a structural model, explicitly designed for sampling in tick time, to correct for the microstructure noise. This section discusses a recently introduced approach by Hansen and Lunde (2004b) that relies on a model free correction of the microstructure noise. The correction is based on the autocorrelation induced by the microstructure noise. Here we define their measure for  $RV$  and compare the results of this measure to the results of our structural approach in section 5.5.

Hansen and Lunde (2004b) start with the general notion that prices are contaminated by noise, which introduces autocorrelations in observed returns. The correction they propose assumes that  $RV$  is biased similar to (5.18). However, their approach relies on calendar time sampling instead of tick time sampling. When sampling in calendar time using the previous tick rule, Hansen and Lunde (2004b) prove that this bias can be corrected for by correcting for all induced higher order autocorrelations in the observed returns. The correction they propose reads

$$RV_{C,q,t}^{(h)} = \sum_{i=1}^{h^{-1}} \{r_{C,t-1+ih}^{(h)}\}^2 + 2 \sum_{s=1}^q \frac{h^{-1}}{h^{-1}-s} \sum_{i=1}^{h^{-1}-s} r_{C,t-1+ih}^{(h)} r_{C,t-1+(i+s)h}^{(h)}, \quad (5.20)$$

where  $q$  is the maximum number of lags, which depends on the sampling frequency  $h$ .

There are several important findings in their paper. Firstly, the best unbiased measure for  $RV$  is achieved when sampling at the highest frequency (in their case 1 second). Second, their correction is not done over a fixed number of lags but depends on a certain amount of time that has to be corrected for. By this we mean that autocorrelation has to be corrected for e.g. 60 seconds, independent of the sampling frequency adopted. Further, the lag length chosen differs per asset.

The correction over longer lags is in contrast with the notion that the microstructure noise induces an MA(1) in observed returns, but can be explained by the time scale that the process evolves in. As mentioned in section 5.2 sampling calendar time at the highest possible frequency using the previous tick rule, leads to many zero observations. Although

the price process can be assumed to evolve continuously, the microstructure noise is only observed when a new transaction occurs. Therefore, if the average time between transactions is 10 seconds, most of the negative autocorrelation will be observed around the 10<sup>th</sup> lag. The autocorrelation function will therefore display negative autocorrelations up to longer lags and will be influenced by the distribution of durations between trades.

If only the time scale explain this autocorrelation function for observed calendar time returns, then we can also address the Hansen and Lunde (2004b) correction in tick time. In a similar fashion we define  $RV$  in tick time as

$$RV_{T,q,t}^{(a)} = \sum_{j=1}^{k(t)^{-1}} \{r_{T,t-1+jak(t)}^{(a)}\}^2 + 2 \sum_{s=1}^q \frac{k(t)^{-1}}{k(t)^{-1} - s} \sum_{j=1}^{k(t)^{-1}-s} r_{T,t-1+jak(t)}^{(a)} r_{T,t-1+(j-s)ak(t)}^{(a)}, \quad (5.21)$$

where  $q$  is the maximum number of lags in tick time, which depends on the sampling frequency ( $a$ ). When sampling is done at every transaction, we expect only a first order negative autocorrelation, induced by the microstructure noise. Applying this correction in tick time should therefore reveal whether there are other effects present in observed returns.

## 5.4 Data

In this study we examine transaction data of 20 actively traded stocks at Nasdaq. The selected stocks vary in liquidity, but were all included in the Nasdaq-100 index during 1999. Data were provided by Nastraq. The sample period extends from February 1<sup>st</sup> until July 30<sup>th</sup>, spanning a total of 124 trading days and contains all transactions within normal trading hours (9.30 - 16.00) of the Nasdaq National Market. The transaction data contains the reported time of the trade, the executed time of the trade, the price at which the trade took place and the volume traded. Some indicators were added to trades that were reported late.

As we are interested in evaluating informative transactions, we remove all trades with a late reported indicator. In a similar way we remove all trades that were executed before, but reported at the same time or after the next executed trade. If two trades were executed at the same time, the trade reported first is considered and the other one is removed. This procedure removes a substantial amount of the transaction data. If these transactions are not removed from the data then tick data would be noisy as we constantly observe past information. The data set we construct contains all trades where the discrepancy between executed and reported time is minimized and thus has the highest information content.

From this data set we remove all outliers. Observations are considered outliers when there are either two consecutive observations that deviate more than 5 standard deviations from the mean, or a single observation that deviates more than 10 standard deviations. The information used from this data set are the log price at which the transaction takes place

Table 5.1: List of Company Ticker symbols and Company names

Ticker Symbol	Company name	Total Trades	Mean(Obs)	Min	Max	Average Duration	Max Duration
AAPL	Apple Computer Inc.	229,790	2933	1205	9919	9.68	549
AMAT	Applied Materials Inc.	651,363	6714	3893	12914	4.45	419
AMGN	Amgen Inc.	387,701	3815	1697	9061	7.48	579
AMZN	Amazon.com, Inc.	893,946	9793	4732	17145	3.25	182
ATHM	At Home Corporation	641,855	6760	2332	17134	4.52	604
CMGI	CMGI Inc.	787,951	8462	2588	17809	3.68	339
COMS	3Com Corporation	529,780	5598	2296	16907	5.48	485
CPWR	Compuware Corporation	314,099	3044	1304	12596	9.24	543
CSCO	Cisco Systems Inc.	1,006,795	11233	6831	18241	2.88	225
DELL	Dell Computer Corporation	1,218,852	13842	7808	22121	2.38	170
INTC	Intel Corp.	1,024,335	11442	7066	17870	2.83	523
MSFT	Microsoft Corp.	1,099,617	12672	7637	19342	2.64	195
NOVL	Novell Inc.	298,136	2940	1371	11436	9.73	572
NXTL	Nextel Communications CL-A	331,468	3284	1257	11430	8.75	614
ORCL	Oracle Corporation	771,807	8215	4815	18463	3.76	511
PSFT	Peoplesoft Inc.	262,246	2596	1327	8194	11.07	591
QWST	Qwest Communications Intl Inc.	497,824	5096	1687	18939	5.82	592
SBUX	Starbucks Corporation	213,749	2051	639	18854	13.58	637
SUNW	Sun Microsystems Inc.	721,561	7743	4361	14459	4.02	475
WCOM	MCI WorldCom Inc.	698,576	7493	4551	17571	4.15	345

This table reports summary statistics of 20 Nasdaq stocks sampled during the period February 1<sup>st</sup> 1999 until July 30<sup>th</sup> 1999. The first two columns report the ticker symbols and the names of the companies. The next column reports the total number of trades over the 124 trading days. The next two columns report the minimum and the maximum number of transactions that occur per day. The last two columns report the average duration and maximum duration between transactions in seconds.



$(p_{t+jk(t)})$  and the reported time of the  $j^{th}$  transaction  $\ell_{t+jk(t)}$ .

Table 5.1 reports some summary statistics of the data. In our filtered sample, the most frequently traded stock, Dell computer corporation, had 1,218,852 transactions over 124 trading days. This averages to 13,842 trades per day. In contrast the least liquid stock in the sample had 213,749 trades over the whole period, resulting in 2,051 trades per day on average. More interesting are the large differences between the minimum and the maximum number of transactions that are observed within a day. This clearly indicates that, there is a great variety in the liquidity of these stocks, although all these stocks were included in the Nasdaq-100 index. The diversity in liquidity is also confirmed by the average duration between transactions. For the most liquid stocks in the sample this duration is below 15 seconds, whereas for the least liquid stock in the sample this lies around one minute. The maximum duration between transactions illustrates that the distribution of durations is right skewed (see Engle and Russell (1997)).

We also sample the data at 2 ticks, 5 ticks and 10 ticks. We consider these data aggregated to analyze the effect of data aggregation on realized volatility. To compare our results to the traditional method of computing realized volatility, we also construct a price series sampled at 1, 10, 30, 60, 120, 300, 600 and 1800 second intervals, using the previous-tick rule.

## 5.5 Results

In this section we present the results. Instead of reporting Realized Variance we report the square root of this, also called the realized volatility ( $rv$ ). Further, instead of using the superscripts ( $h$ ) and ( $k(t)$ ) we will use the sampling frequency as the superscript, when addressing  $rv$  in the tables. We start by looking at  $rv$  at different sampling frequencies, when no correction for the microstructure noise has been done. Next, we address the parameter estimates of the structural model introduced in section 5.3.2, for different sampling frequencies and report the realized volatilities found by this approach. Finally, we report realized volatilities when sampling is done in both calendar time and tick time, using the Hansen and Lunde (2004b) corrections discussed in section 5.3.3.

### 5.5.1 Uncorrected Realized Volatility

In this subsection we discuss the results for the realized volatility, computed at different sampling frequencies, when no correction has been done for the microstructure noise present. We only report the results for sampling in calendar time as tick time sampling leads to a similar pattern in  $rv$ . The realized volatilities have been computed at different sampling frequencies.

In table 5.2 we report the results for average uncorrected realized volatilities at a 1,

Table 5.2: Realized Volatilities at different Sampling Frequencies

	1 sec	10 sec	30 sec	60 sec	120 sec	300 sec	600 sec	1800 sec
AAPL	6.15 (1.34)	4.78 (0.88)	3.86 (0.65)	3.40 (0.59)	3.11 (0.63)	2.87 (0.72)	2.73 (0.77)	2.66 (1.00)
AMAT	6.62 (1.32)	4.47 (0.83)	3.66 (0.78)	3.50 (0.84)	3.45 (0.89)	3.36 (0.98)	3.36 (1.05)	3.37 (1.19)
AMGN	5.98 (1.44)	4.47 (0.86)	3.47 (0.66)	3.02 (0.65)	2.81 (0.70)	2.73 (0.89)	2.60 (0.92)	2.48 (1.11)
AMZN	9.79 (2.35)	6.02 (1.32)	5.09 (1.21)	4.98 (1.27)	4.92 (1.32)	4.95 (1.43)	4.83 (1.44)	4.60 (1.55)
ATHM	10.33 (3.03)	6.77 (1.64)	5.41 (1.46)	5.05 (1.52)	4.93 (1.63)	4.84 (1.83)	4.70 (1.90)	4.54 (2.00)
CMGI	11.54 (3.89)	7.31 (2.24)	6.17 (2.07)	5.96 (2.16)	5.85 (2.27)	5.83 (2.49)	5.61 (2.62)	5.30 (2.92)
COMS	10.42 (2.72)	7.15 (1.18)	5.07 (0.75)	4.11 (0.76)	3.53 (0.89)	3.10 (1.18)	2.89 (1.39)	2.76 (1.65)
CPWR	9.75 (3.40)	7.35 (2.14)	5.74 (1.48)	4.96 (1.28)	4.44 (1.29)	4.07 (1.40)	3.89 (1.58)	3.80 (1.93)
CSCO	5.64 (1.03)	3.26 (0.44)	2.55 (0.41)	2.42 (0.47)	2.37 (0.52)	2.33 (0.60)	2.27 (0.64)	2.23 (0.81)
DELL	10.40 (2.45)	5.43 (1.16)	3.78 (0.80)	3.33 (0.71)	3.11 (0.75)	2.99 (0.93)	2.88 (1.01)	2.78 (1.12)
INTC	6.45 (1.68)	3.63 (0.78)	2.69 (0.50)	2.49 (0.47)	2.43 (0.52)	2.38 (0.59)	2.36 (0.63)	2.28 (0.84)
MSFT	5.23 (1.17)	2.92 (0.50)	2.24 (0.40)	2.10 (0.42)	2.07 (0.48)	2.05 (0.54)	1.99 (0.58)	2.01 (0.66)
NOVL	8.55 (1.86)	6.65 (1.06)	5.19 (0.72)	4.34 (0.64)	3.72 (0.65)	3.26 (0.75)	3.04 (0.81)	2.77 (0.90)
NXTL	7.11 (2.06)	5.39 (1.24)	4.20 (0.93)	3.62 (0.88)	3.21 (0.96)	2.92 (1.02)	2.77 (1.10)	2.56 (1.18)
ORCL	11.04 (3.25)	6.69 (1.46)	4.60 (0.88)	3.89 (0.81)	3.51 (0.85)	3.22 (1.01)	3.08 (1.10)	3.03 (1.37)
PSFT	12.29 (3.06)	9.73 (1.97)	7.59 (1.36)	6.22 (1.09)	5.16 (1.09)	4.16 (1.15)	3.69 (1.20)	3.25 (1.35)
QWST	8.37 (2.64)	5.84 (1.31)	4.41 (1.00)	3.89 (1.05)	3.65 (1.20)	3.43 (1.35)	3.28 (1.38)	3.26 (1.62)
SBUX	7.13 (2.43)	5.77 (1.50)	4.80 (1.13)	4.16 (0.96)	3.67 (0.90)	3.21 (0.96)	2.95 (1.03)	2.77 (1.14)
SUNW	6.30 (1.71)	4.13 (0.91)	3.23 (0.72)	3.05 (0.74)	2.98 (0.80)	2.90 (0.91)	2.84 (1.00)	2.92 (1.24)
WCOM	4.82 (0.84)	3.16 (0.49)	2.33 (0.43)	2.11 (0.48)	2.05 (0.54)	1.95 (0.58)	1.88 (0.62)	1.81 (0.71)

Note: This Table reports average uncorrected realized volatilities ( $\overline{rv}_C^{(h)}$ ) sampled at different calendar time frequencies  $h$ . These realized volatilities are computed as

$$\overline{rv}_C^{(h)} = \frac{1}{N} \sum_{t=1}^N \sqrt{\sum_{i=1}^{h-1} \{r_{C,t-1+i\Delta}^{(h)}\}^2},$$

where  $N = 124$  is the total number of trading days. Standard deviations of  $\overline{rv}_C^{(h)}$  over the 124 days are reported in brackets.

10, 30, 60, 120, 300, 600 and 1800 second sampling frequency and represent the so called volatility signature of a stock. The table reports the average realized volatility over the 124 trading days and reports the standard deviations of  $rv$  in brackets. These results confirm the general notion that at very high sampling frequencies microstructure noise contaminates the observed prices, which leads to an upward bias for  $rv$ . At lower frequencies  $rv$  decreases. Similar results have been documented by Barndorff-Nielsen and Shephard (2002) and Hansen and Lunde (2004b).

There are some interesting features observed from table 5.2. Firstly, we find that for most stocks the lowest value for  $rv$  is found at the 30 minute (1800 sec) sampling frequency. This even holds for the most liquid stocks in the sample like Dell and Intel. This suggests that when sampling at e.g. a five minute frequency, a frequency often adopted, there is still a considerable amount of microstructure noise present in the measure of  $rv$ . It further shows that the effects of microstructure are persistent in aggregated returns and even affect lower frequency returns.

Another interesting feature of table 5.2 is that the lowest standard deviations for average  $rv$  are found around the 30 and 60 second sampling frequency. When sampling at higher frequencies the bias in  $rv$  due to the microstructure noise likely causes daily  $rv$ 's to be more volatile. At lower frequencies the increase in the standard deviations of  $rv$  can be explained by the measurement error introduced by sampling at lower frequencies (see Barndorff-Nielsen and Shephard (2002)).

Hansen and Lunde (2004b) use the 30 minute  $rv$  as a benchmark volatility, to which they compare the realized volatilities computed with bias corrections. In the remainder of the discussion we will use this level of volatility as a benchmark.

The next section addresses the results of a structural model, designed to correct for the noise in  $rv$ .

### 5.5.2 Structural Model: Parameter Estimates

In this subsection we discuss the parameter estimates of the model developed in section 5.2. The parameters are calibrated using the full sample and are not re-estimated on a daily basis. Therefore the subscript  $t$  for the parameters is omitted. We estimate this model at three different sampling frequencies, the 1-tick, 2-tick and 5-tick level.<sup>5</sup> In tables 5.3 - 5.5 we report the results of these parameters.

The results of the model when sampled at every tick are reported in table 5.3. The first column reports the duration parameter ( $\delta_1$ ) on the innovation of the efficient price. The two values of interest for this parameter are  $\frac{1}{2}$  and 0. When  $\delta_1 = \frac{1}{2}$  the random walk evolves

<sup>5</sup>Higher aggregation levels are not considered as the Kalman Filter routine uses 50 observations at the start of the day to initialize the system. At the higher aggregation level this leads to a loss of data and does not incorporate the first part of the day in the estimation of the parameters.

Table 5.3: Parameter Estimates sampled at 1-tick

Stock	$\delta_1$	$\delta_2$	$\alpha$	$\sigma$	$\omega_1$	$\omega_2$	$\omega_3$
AAPL	0.10 (0.01)	-0.03 (0.01)	2.58 (0.04)	4.90 (0.03)	4.91 (0.10)	0.00 (0.00)	2.68 (0.16)
AMAT	0.08 (0.01)	-0.06 (0.01)	2.34 (0.03)	3.61 (0.01)	3.78 (0.06)	0.38 (0.21)	2.16 (0.08)
AMGN	0.05 (0.01)	0.03 (0.05)	2.37 (1.46)	3.82 (0.02)	5.72 (1.86)	3.36 (3.17)	4.11 (2.59)
AMZN	0.00 (0.01)	0.04 (0.01)	1.83 (0.10)	4.52 (0.02)	6.98 (0.16)	1.50 (0.72)	4.70 (0.23)
ATHM	0.08 (0.01)	-0.06 (0.01)	1.80 (1.09)	5.54 (0.02)	8.88 (1.87)	3.92 (4.26)	6.13 (2.71)
CMGI	0.11 (0.01)	-0.08 (0.01)	1.79 (0.14)	6.13 (0.02)	8.62 (0.29)	1.27 (1.91)	5.08 (0.48)
COMS	-0.07 (0.01)	0.28 (0.04)	1.96 (0.39)	3.68 (0.02)	11.00 (0.24)	9.67 (0.27)	10.08 (0.25)
CPWR	0.00 (0.01)	0.06 (0.03)	1.94 (0.74)	6.72 (0.06)	12.62 (1.30)	8.50 (1.94)	9.24 (1.81)
CSCO	0.03 (0.01)	0.02 (0.01)	1.87 (0.33)	2.04 (0.01)	3.73 (0.18)	2.85 (0.24)	3.12 (0.22)
DELL	-0.17 (0.03)	0.36 (0.05)	1.85 (0.45)	2.25 (0.02)	8.17 (0.14)	6.23 (0.17)	6.50 (0.16)
INTC	0.04 (0.01)	0.02 (0.01)	1.82 (0.08)	2.01 (0.01)	4.72 (0.04)	3.81 (0.04)	4.07 (0.04)
MSFT	0.03 (0.01)	0.02 (0.01)	1.83 (0.41)	1.66 (0.01)	3.61 (0.15)	2.77 (0.20)	3.08 (0.18)
NOVL	0.16 (0.01)	-0.04 (0.06)	2.02 (0.87)	5.88 (0.05)	9.71 (1.55)	7.93 (1.90)	8.78 (1.71)
NXTL	-0.01 (0.01)	0.12 (0.09)	2.02 (1.50)	4.51 (0.03)	8.37 (1.81)	6.28 (2.40)	7.33 (2.06)
ORCL	-0.06 (0.01)	0.13 (0.02)	2.02 (0.50)	3.16 (0.02)	9.77 (0.26)	8.60 (0.29)	8.81 (0.29)
PSFT	0.05 (0.01)	0.00 (0.01)	5.42 (0.14)	7.86 (0.08)	8.91 (0.26)	0.00 (0.00)	2.25 (0.94)
QWST	0.04 (0.01)	-0.03 (0.01)	2.01 (0.04)	4.39 (0.02)	8.42 (0.09)	5.77 (0.11)	6.51 (0.11)
SBUX	0.12 (0.01)	-0.06 (0.02)	2.01 (0.57)	6.71 (0.06)	9.49 (1.38)	5.85 (2.24)	7.50 (1.72)
SUNW	0.04 (0.01)	-0.01 (0.01)	2.01 (0.12)	2.99 (0.01)	4.54 (0.12)	3.09 (0.18)	3.64 (0.15)
WCOW	0.10 (0.01)	-0.05 (0.04)	2.03 (1.72)	2.13 (0.01)	3.64 (1.07)	2.76 (1.41)	3.07 (1.26)

Note: This table reports the parameter estimates with standard errors in brackets, of the model

$$p_{t+jk(t)} = p_{t+jk(t)}^* + u_{t+jk(t)},$$

$$p_{t+jk(t)}^* = p_{t+(j-1)k(t)}^* + \tau_{t+jk(t)}^{\delta_1} \varepsilon_{t+jk(t)},$$

$$u_{t+jk(t)} = \alpha \tau_{t+jk(t)}^{\delta_2} \varepsilon_{t+jk(t)} + e_{1,t+jk(t)} + e_{2,t+jk(t)} + e_{3,t+jk(t)},$$

where  $\text{Var}(\varepsilon_{t+jk(t)}) = \sigma^2$  and  $\text{Var}(e_{1,t+jk(t)}) = \omega_1^2$ ,  $\text{Var}(e_{2,t+jk(t)}) = \omega_2^2$ ,  $\text{Var}(e_{3,t+jk(t)}) = \omega_3^2$ . The sampling frequency is at the 1 tick level. The first column reports the estimates for the duration parameter on the random walk innovation  $\delta_1$ . The second column reports the parameter estimates for the time impact on the correlation between the efficient price and the microstructure noise. The next column reports the  $\alpha$  in the model which measures the correlation between efficient price innovation and microstructure noise at average durations. The next column reports the volatility of the efficient price for a single transaction at average durations. The last three columns report the standard deviation of the microstructure noise, where  $\omega_1$  measures the noise near the opening (9.30 - 11.00),  $\omega_2$  in the middle of the trading day (11.00 - 14.30) and  $\omega_3$  near the close of the market (14.30 - 16.00). The standard deviations in the last 4 columns are multiplied by 100.

in calendar time and the variance of the efficient price expands proportional to the time interval between transactions. When  $\delta_1 = 0$  the random walk evolves in transaction time. This time scale has more recently been motivated by Ané and Geman (2000). The results for  $\delta_1$  differ substantially over the stocks in the sample but are on average positive with a value of 0.04. For 14 stocks this value is significantly larger than 0 and only 3 stocks report a significantly negative value. The parameter estimates are in all cases significantly different from  $\frac{1}{2}$  indicating that the hypothesis of the price process evolving in calendar time is clearly rejected. These results are in line with the findings of Frijns and Schotman (2004).

Consequently, we are also interested in the impact of time on the correlation between the microstructure noise and the innovation in the efficient price, measure by the sum of  $\delta_1$  and

$\delta_2$ . The second column of table 5.3 reports the parameter estimates of  $\delta_2$ . Again there is a large difference between the parameter values for the individual stocks. The average of these values is 0.03, however there are 6 stocks for which the parameter is significantly negative compared to 8 significant positive values. We typically find that  $\delta_2$  takes the opposite sign of the parameter  $\delta_1$  in all but three cases. In combination with  $\delta_1$  this reduces the effect of time on the correlation (see Equation (5.18)). However, for all stocks the impact of time on the correlation is positive, which refers to higher correlations at longer durations.

The third column of table 5.3 reports the  $\alpha$  of the model. This parameter measures the extend to which the innovation in the random walk correlates with the microstructure noise. A first conclusion from these parameters is that all values for  $\alpha$  are positive, although not all parameter values are significant. The average value for  $\alpha$  is 2.18. Values of  $\alpha$  larger than 2 indicate that the noise around the efficient price is more than twice the innovation in the random walk. One point to consider is that whenever  $\delta_2$  is not significantly different from zero, the identification of  $\alpha$  is not guaranteed and vice versa.

The next column reports the results on the innovation parameter of the random walk ( $\sigma$ ), expressed as the square root of the variance of the innovation in the efficient price. These parameters are estimated with very high precision as indicated by their low standard errors. The most interesting result for these parameters is that the innovation in the efficient price is inversely related to the liquidity of the asset. For the most liquid stocks the innovation in the random walk per transaction is smaller. For the illiquid stocks the innovation is larger. Therefore, there seems to be a liquidity effect in the innovation in the efficient price, but no relationship with the time between transactions.

The last three columns in the table report the microstructure noise around the efficient price. The parameter  $\omega_1$  measures the noise near the open of the market (all trades between 9.30 and 11.00),  $\omega_2$  measures the noise during the day (11.00 until 14.30) and  $\omega_3$  measures the noise around the close (from 14.30 - 16.00). Overall we observe that the noise is highest near the opening of the market, decreases and is lowest in the middle of the trading day and increases again at the close of the day. These results are in line with Madhavan, Richardson, and Roomans (1997). Again we observe the same relationship as with the innovation in the random walk. The more liquid the stock, the lower the microstructure noise.

For the aggregated data at the 2-tick level and the 5-tick level we discuss the results simultaneously. A first overall result that we observe is that the  $\delta_1$  parameter decreases, where at the 5-tick level we observe only significant negative values. Where the value for  $\delta_1$  was 0.04 for the 1-tick level, we now have -0.05 and -0.15 for the 2- and 5-tick level respectively. These results indicate that when for the aggregate data the duration between multiple transactions is short, the innovation in the efficient price is larger and vice versa.

The opposite result is observed for the  $\delta_2$  parameter. When data is aggregated this term increases. Where the average value for this parameter was 0.03 for the 1-tick level, this

Table 5.4: Parameter Estimates sampled at 2-tick

Stock	$\delta_1$	$\delta_2$	$\alpha$	$\sigma$	$\omega_1$	$\omega_2$	$\omega_3$
AAPL	0.03 (0.01)	0.10 (0.01)	1.32 (0.03)	6.70 (0.04)	4.95 (0.14)	0.00 (0.01)	2.91 (0.21)
AMAT	0.01 (0.01)	0.01 (0.01)	1.02 (0.01)	5.22 (0.02)	3.74 (0.07)	0.37 (0.20)	2.28 (0.11)
AMGN	-0.03 (0.01)	0.17 (0.13)	1.07 (0.98)	5.32 (0.03)	5.74 (2.39)	3.67 (3.76)	4.46 (3.09)
AMZN	-0.04 (0.01)	0.07 (0.02)	0.78 (0.01)	6.66 (0.03)	6.75 (0.07)	0.11 (0.04)	4.42 (0.08)
ATHM	0.03 (0.01)	-0.03 (0.02)	0.77 (0.18)	7.97 (0.04)	8.76 (0.62)	3.98 (1.37)	6.11 (0.88)
CMGI	0.07 (0.01)	-0.07 (0.03)	0.72 (0.60)	8.95 (0.04)	8.50 (2.81)	1.64 (14.63)	5.51 (4.34)
COMS	-0.19 (0.01)	0.51 (0.03)	1.01 (0.11)	4.93 (0.03)	11.06 (0.13)	9.81 (0.14)	10.18 (0.13)
CPWR	-0.07 (0.01)	0.16 (0.02)	0.97 (0.25)	9.02 (0.07)	12.86 (0.79)	8.88 (1.15)	9.49 (1.07)
CSCO	-0.06 (0.01)	0.17 (0.11)	0.78 (0.74)	2.96 (0.01)	3.67 (0.88)	2.89 (1.12)	3.15 (1.02)
DELL	-0.28 (0.02)	0.77 (0.05)	0.81 (0.16)	3.20 (0.02)	8.06 (0.13)	6.21 (0.13)	6.59 (0.13)
INTC	-0.07 (0.01)	0.32 (0.02)	0.65 (0.06)	2.90 (0.01)	4.80 (0.05)	3.97 (0.06)	4.20 (0.06)
MSFT	-0.08 (0.01)	0.23 (0.02)	0.62 (0.13)	2.39 (0.01)	3.68 (0.10)	2.92 (0.13)	3.26 (0.12)
NOVL	0.06 (0.01)	0.05 (0.01)	2.29 (0.05)	7.66 (0.05)	5.41 (0.23)	0.00 (0.00)	3.88 (0.29)
NXTL	-0.11 (0.01)	0.21 (0.04)	2.15 (0.76)	6.02 (0.05)	5.25 (2.58)	1.76 (7.67)	3.98 (3.40)
ORCL	-0.17 (0.01)	0.21 (0.02)	2.18 (0.22)	4.39 (0.02)	8.24 (0.27)	7.18 (0.31)	7.43 (0.30)
PSFT	-0.08 (0.01)	0.17 (0.01)	3.66 (0.11)	9.73 (0.11)	8.71 (0.39)	0.00 (0.00)	2.17 (1.45)
QWST	-0.01 (0.01)	0.00 (0.01)	1.68 (0.51)	6.23 (0.03)	6.20 (1.65)	1.68 (6.00)	3.68 (2.76)
SBUX	0.05 (0.01)	0.02 (0.01)	1.62 (0.04)	8.83 (0.07)	7.33 (0.26)	0.15 (0.14)	4.40 (0.34)
SUNW	-0.04 (0.01)	0.08 (0.01)	1.32 (0.76)	4.30 (0.02)	3.39 (2.07)	1.37 (5.15)	2.33 (3.01)
WCOM	-0.01 (0.01)	0.11 (0.07)	1.36 (0.88)	3.02 (0.01)	3.06 (1.30)	1.87 (2.13)	2.36 (1.69)

Note: This table reports the parameter estimates with standard errors in brackets, of the model

$$p_{t+jk(t)} = p_{t+jk(t)}^* + u_{t+jk(t)},$$

$$p_{t+jk(t)}^* = p_{t+(j-1)k(t)}^* + \tau_{t+jk(t)}^{\delta_1} \varepsilon_{t+jk(t)},$$

$$u_{t+jk(t)} = \alpha \tau_{t+jk(t)}^{\delta_2} \varepsilon_{t+jk(t)} + e_{1,t+jk(t)} + e_{2,t+jk(t)} + e_{3,t+jk(t)}.$$

where  $\text{Var}(\varepsilon_{t+jk(t)}) = \sigma^2$  and  $\text{Var}(e_{1,t+jk(t)}) = \omega_1^2$ ,  $\text{Var}(e_{2,t+jk(t)}) = \omega_2^2$ ,  $\text{Var}(e_{3,t+jk(t)}) = \omega_3^2$ . The sampling frequency is at the 2 tick level. The first column reports the estimates for the duration parameter on the random walk innovation  $\delta_1$ . The second column reports the parameter estimates for the time impact on the correlation between the efficient price and the microstructure noise. The next column reports the  $\alpha$  in the model which measures the correlation between efficient price innovation and microstructure noise at average durations. The next column reports the volatility of the efficient price for a single transaction at average durations. The last three columns report the standard deviation of the microstructure noise, where  $\omega_1$  measures the noise near the opening (9.30 - 11.00),  $\omega_2$  in the middle of the trading day (11.00 - 14.30) and  $\omega_3$  near the close of the market (14.30 - 16.00). The standard deviations in the last 4 columns are multiplied by 100.

increases to 0.16 and 0.51 for the 2- and the 5-tick level respectively. The impact of time on the correlation between the innovation in the efficient price and the microstructure noise, defined as the sum of  $\delta_1$  and  $\delta_2$ , tends to be larger when durations are longer than average. Again the overall result holds in each case that  $\delta_2$  take the opposite sign of  $\delta_1$ , but the sum remains positive.

For the  $\alpha$ 's we see a decreasing trend. The average of 2.18 for the 1-tick level, decreases to 1.34 and 0.58 for the 2- and the 5-tick level respectively.

An interesting result comes from the idiosyncratic noise component. This noise does not change at higher aggregation levels. However, as the innovation term of the efficient price becomes larger when data is aggregated, the signal to noise ratio increases. Therefore, the

Table 5.5: Parameter Estimates sampled at 5-tick

Stock	$\delta_1$	$\delta_2$	$\alpha$	$\sigma$	$\omega_1$	$\omega_2$	$\omega_3$
AAPL	-0.06 (0.01)	0.23 (0.02)	0.49 (0.02)	10.51 (0.09)	3.97 (0.35)	0.00 (0.00)	2.41 (0.47)
AMAT	-0.08 (0.01)	0.22 (0.03)	0.27 (0.01)	8.68 (0.04)	3.18 (0.15)	0.01 (0.00)	2.12 (0.19)
AMGN	-0.13 (0.01)	0.32 (0.04)	0.42 (0.12)	8.57 (0.07)	4.88 (0.89)	2.89 (1.52)	3.71 (1.19)
AMZN	-0.11 (0.01)	0.32 (0.06)	0.15 (0.01)	11.36 (0.06)	6.76 (0.19)	0.00 (0.00)	5.04 (0.14)
ATHM	-0.07 (0.01)	1.69 (0.14)	0.01 (0.00)	13.16 (0.08)	10.09 (0.16)	6.57 (0.12)	8.33 (0.14)
CMGI	-0.02 (0.01)	1.35 (0.17)	0.02 (0.01)	15.06 (0.08)	9.76 (0.17)	5.12 (0.19)	7.74 (0.17)
COMS	-0.32 (0.01)	0.45 (0.02)	2.10 (0.05)	7.62 (0.06)	4.58 (0.28)	0.09 (0.02)	2.31 (0.47)
CPWR	-0.16 (0.01)	0.20 (0.02)	0.76 (0.03)	13.92 (0.14)	8.56 (0.48)	1.27 (1.25)	4.32 (0.74)
CSCO	-0.19 (0.01)	0.36 (0.03)	0.50 (0.02)	4.93 (0.02)	2.22 (0.11)	0.55 (0.34)	1.54 (0.15)
DELL	-0.40 (0.02)	0.66 (0.03)	1.56 (0.02)	5.30 (0.03)	4.95 (0.20)	0.00 (0.00)	2.11 (0.14)
INTC	-0.19 (0.01)	0.39 (0.04)	0.73 (0.12)	4.89 (0.02)	2.65 (0.54)	0.83 (1.69)	1.74 (0.82)
MSFT	-0.21 (0.01)	0.37 (0.03)	0.63 (0.05)	4.02 (0.02)	2.16 (0.19)	0.44 (0.85)	1.44 (0.27)
NOVL	-0.05 (0.01)	0.23 (0.02)	0.99 (0.03)	11.56 (0.10)	4.52 (0.54)	0.00 (0.00)	3.50 (0.58)
NXTL	-0.23 (0.01)	0.36 (0.07)	0.84 (0.43)	9.40 (0.09)	3.90 (4.91)	1.46 (13.17)	3.89 (4.95)
ORCL	-0.30 (0.01)	0.38 (0.01)	0.85 (0.26)	7.08 (0.04)	7.53 (0.87)	6.96 (0.96)	7.12 (0.93)
PSFT	-0.22 (0.01)	0.34 (0.02)	1.78 (0.07)	14.06 (0.18)	6.10 (0.97)	0.00 (0.00)	1.05 (4.80)
QWST	-0.08 (0.01)	1.29 (2.54)	0.00 (0.02)	10.14 (0.07)	9.46 (0.14)	7.73 (0.14)	8.33 (0.14)
SBUX	0.02 (0.01)	1.28 (0.94)	0.00 (0.00)	13.82 (0.16)	12.50 (0.28)	10.41 (0.16)	11.61 (0.22)
SUNW	-0.15 (0.01)	-0.04 (0.69)	-0.17 (0.60)	7.19 (0.04)	6.22 (2.49)	5.44 (2.85)	5.77 (2.69)
WCOM	-0.15 (0.01)	-0.19 (0.24)	-0.35 (0.26)	4.98 (0.03)	5.18 (0.62)	4.73 (0.68)	4.96 (0.65)

Note: This table reports the parameter estimates with standard errors in brackets, of the model

$$\begin{aligned}
p_{t+jk(t)} &= p_{t+jk(t)}^* + u_{t+jk(t)}, \\
p_{t+jk(t)}^* &= p_{t+(j-1)k(t)}^* + \tau_{t+jk(t)}^{\delta_1} \varepsilon_{t+jk(t)}, \\
u_{t+jk(t)} &= \alpha \tau_{t+jk(t)}^{\delta_2} \varepsilon_{t+jk(t)} + e_{1,t+jk(t)} + e_{2,t+jk(t)} + e_{3,t+jk(t)}.
\end{aligned}$$

where  $\text{Var}(\varepsilon_{t+jk(t)}) = \sigma^2$  and  $\text{Var}(e_{1,t+jk(t)}) = \omega_1^2$ ,  $\text{Var}(e_{2,t+jk(t)}) = \omega_2^2$ ,  $\text{Var}(e_{3,t+jk(t)}) = \omega_3^2$ . The sampling frequency is at the 5 tick level. The first column reports the estimates for the duration parameter on the random walk innovation  $\delta_1$ . The second column reports the parameter estimates for the time impact on the correlation between the efficient price and the microstructure noise. The next column reports the  $\alpha$  in the model which measures the correlation between efficient price innovation and microstructure noise at average durations. The next column reports the volatility of the efficient price for a single transaction at average durations. The last three columns report the standard deviation of the microstructure noise, where  $\omega_1$  measures the noise near the opening (9.30 - 11.00),  $\omega_2$  in the middle of the trading day (11.00 - 14.30) and  $\omega_3$  near the close of the market (14.30 - 16.00). The standard deviations in the last 4 columns are multiplied by 100.

total impact of microstructure noise disappears at higher aggregation levels, as expected.

### 5.5.3 Structural Model: Realized Volatility

In this subsection, we discuss the results of realized volatility given the structural model proposed in section 5.3.2. On each of the 124 trading days we calculate  $rv$  based on the parameter estimates discussed in the previous section. These  $rv$ 's are computed using a 1, 2 and 5 tick sampling frequency. Additionally, we compute  $rv$  by estimating (5.17) on a daily basis. This is only done for the 1 tick sampling frequency.<sup>6</sup>

<sup>6</sup>We also estimated the model on the lower 2 and 5 tick frequencies. However, due to the lower amount of observations at these frequencies, the parameters for an increasing number of days could not be estimated.

Table 5.6: Realized Volatility using the Kalman Filter

	Panel A			Panel B
	KF 1 tick	KF 2 ticks	KF 5 ticks	Daily KF 1 tick
AAPL	2.25 (0.37)	2.25 (0.43)	2.36 (0.52)	2.21 (0.44)
AMAT	2.53 (0.26)	2.64 (0.30)	2.90 (0.38)	2.49 (0.46)
AMGN	2.07 (0.31)	2.11 (0.36)	2.28 (0.44)	2.02 (0.44)
AMZN	3.82 (0.49)	4.02 (0.54)	4.46 (0.66)	3.71 (0.89)
ATHM	3.80 (0.69)	3.93 (0.78)	4.27 (0.98)	3.70 (1.08)
CMGI	4.65 (0.79)	4.84 (0.89)	5.29 (1.10)	4.36 (1.57)
COMS	2.44 (0.57)	2.48 (0.66)	2.70 (0.80)	2.38 (0.74)
CPWR	3.31 (0.64)	3.30 (0.71)	3.49 (0.84)	3.23 (0.92)
CSCO	1.82 (0.15)	1.91 (0.18)	2.11 (0.23)	1.78 (0.34)
DELL	2.33 (0.29)	2.44 (0.33)	2.69 (0.40)	2.19 (0.50)
INTC	1.81 (0.15)	1.89 (0.19)	2.11 (0.24)	1.75 (0.35)
MSFT	1.55 (0.12)	1.62 (0.14)	1.81 (0.18)	1.52 (0.29)
NOVL	2.65 (0.38)	2.53 (0.45)	2.57 (0.55)	2.64 (0.48)
NXTL	2.29 (0.48)	2.31 (0.55)	2.55 (0.69)	2.25 (0.65)
ORCL	2.53 (0.36)	2.61 (0.42)	2.88 (0.52)	2.45 (0.60)
PSFT	3.46 (0.52)	3.30 (0.61)	3.49 (0.76)	3.37 (0.93)
QWST	2.67 (0.58)	2.73 (0.65)	2.92 (0.77)	2.58 (0.78)
SBUX	2.51 (0.60)	2.40 (0.68)	2.42 (0.73)	2.43 (0.65)
SUNW	2.24 (0.24)	2.35 (0.29)	2.61 (0.36)	2.19 (0.47)
WCOM	1.55 (0.13)	1.60 (0.16)	1.77 (0.22)	1.52 (0.37)

Note: This table reports the Realized volatility for different aggregation levels. Realized volatilities are based on the parameter estimates of equation (5.17) and are computed as

$$rv_{T,p^*,t}^{(a)} = \sqrt{\sum_{j=1}^{k(t)-1} \tau_{t+jak(t)}^{2\delta_1} \sigma^{(a),2}}, \quad rv_{p^*,t}^{(a)} = \sqrt{\sum_{j=1}^{k(t)-1} \tau_{t+jak(t)}^{2\delta_1,t} \sigma_t^{(a),2}},$$

for the full sample and daily calculations, respectively. We report the average values of  $rv$ ,

$$\overline{rv}_{T,p^*}^{(a)} = \frac{1}{N} \sum_{t=1}^N rv_{T,p^*,t}^{(a)},$$

where  $N = 124$  is the total number of trading days. Standard deviations of  $rv$  over the trading days are reported in brackets. Panel A reports the results where parameters are estimated over the full sample, panel B reports the results for the daily estimates.

Panel A in table 5.6 presents the average realized volatility estimates obtained from (5.17) using the transactions data when the Kalman Filter is run over the whole sample period. The standard deviations of these realized volatilities over the 124 trading days are reported in brackets. The results differ substantially from table 5.2. Sampling at the highest frequency results in the lowest average level of realized volatility among the frequencies analyzed. Interestingly, the average levels of realized volatilities are considerably lower compared to the 30 minute  $rv$ 's presented in table 5.2. Moreover the highest sampling frequency also yields the lowest standard deviations for  $rv$ 's. When the model is estimated at a 2 or 5 tick frequency we see that  $rv$  increases.



The results of the model presented above assume fixed parameters over the whole sample period. As a robustness test of our results we relax this restriction and allow the parameters to vary on a daily basis. Panel B in table 5.6 presents the results when the model is estimated on a daily basis at a 1 tick frequency.<sup>7</sup> The results suggest that estimating the model over the whole sample period smooths out some of the daily variations as the standard deviation of average  $rv$  is somewhat higher than for the model estimated over all days. However, the average levels of the realized volatilities are very similar. Our general result that at the highest frequency, realized volatilities are lower are not affected when estimating the model on a daily basis.

As the realized volatilities increase at lower sampling frequencies, we expect some sort of dependence between tick returns. This, however, contradicts the assumption that prices follow a random walk. In the next section we follow a different approach to correcting for the bias in  $rv$ , which might reveal more of the properties of  $rv$ .

### 5.5.4 Model-Free Correction

In this section we present the results for  $rv$  using the Hansen and Lunde (2004b) correction. We start with the correction on the data sampled in calendar time, after which we present the results when sampling is done in tick time.

One of the main findings of Hansen and Lunde (2004b) is that the best measure for  $rv$  is found at the highest sampling frequency. Best in this sense means that the measurement error made between the estimated and true volatility is smallest. The correction to the data, however, is done over longer lags, where Hansen and Lunde (2004b) find the correct lag length when the corrected  $rv$  equals the average 30 minute  $rv$ .

In table 5.7, we report the average realized volatilities sampled in calendar time at every second, where the correction has been done over several lag lengths. Standard deviations of average  $rv$ 's are again reported in brackets. Interestingly, a one lag correction is by far not enough to correct for the impacts of the microstructure noise, which rejects the theoretically assumed MA(1) structure for calendar time returns. The next step is to find the  $rv$  that best corresponds to the 30 minute  $rv$  in table 5.2. We find that the lags needed to correct for the microstructure noise to some extent depends on the liquidity of the asset. For frequently traded stocks like Cisco and Microsoft a five second correction seems to be sufficient. However, for the less liquid ones, such as Starbucks and Compuware, a correction of 5 minutes is required. These findings are all consistent with the findings of Hansen and Lunde (2004b), but there are some other interesting features.

Firstly, when we consider the standard deviations of  $rv$ , we find that the lowest standard deviation is often found at the value for the average  $rv$  that closest corresponds to the 30

---

<sup>7</sup>Numerical issues burden the estimation of the model at lower frequencies.

Table 5.7: Realized Volatility: Calendar time corrected

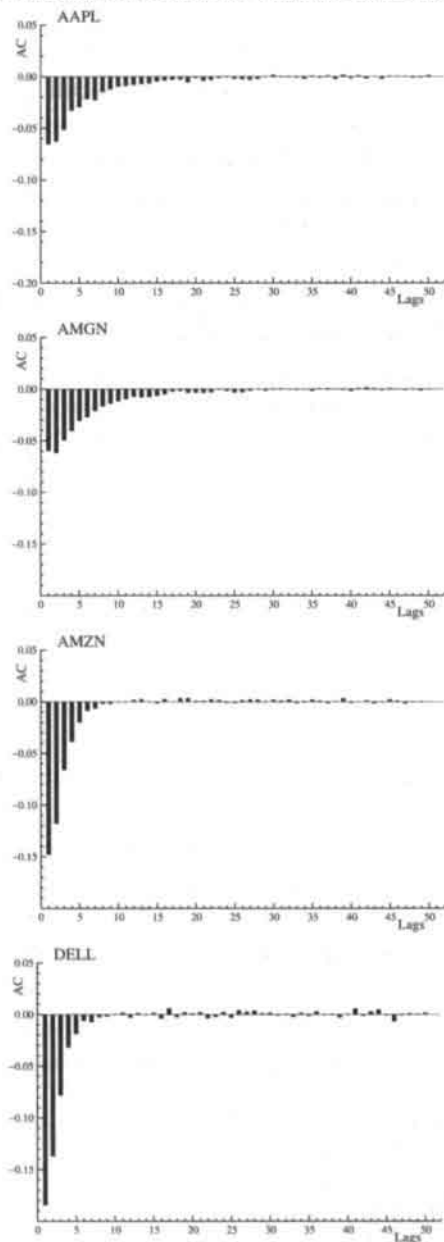
	q = 1	q = 5	q = 10	q = 15	q = 30	q = 45	q = 60	q = 90	q = 120	q = 300
AAPL	5.79 (1.16)	4.54 (0.79)	3.78 (0.63)	3.37 (0.54)	2.92 (0.51)	2.78 (0.58)	2.71 (0.65)	2.69 (0.71)	2.72 (0.73)	2.62 (0.80)
AMAT	5.85 (1.04)	3.71 (0.49)	3.06 (0.46)	2.93 (0.49)	3.09 (0.65)	3.25 (0.73)	3.30 (0.78)	3.35 (0.85)	3.35 (0.85)	3.35 (1.02)
AMGN	5.60 (1.26)	4.19 (0.71)	3.31 (0.55)	2.85 (0.51)	2.45 (0.55)	2.45 (0.67)	2.45 (0.72)	2.55 (0.78)	2.59 (0.83)	2.56 (0.93)
AMZN	8.26 (1.81)	4.70 (0.98)	4.24 (0.95)	4.28 (1.08)	4.65 (1.27)	4.80 (1.31)	4.83 (1.34)	4.91 (1.39)	4.98 (1.46)	4.90 (1.55)
ATHM	9.15 (2.37)	5.71 (1.28)	4.68 (1.25)	4.39 (1.24)	4.50 (1.49)	4.65 (1.57)	4.75 (1.69)	4.82 (1.83)	4.82 (1.85)	4.65 (2.02)
CMGI	9.97 (3.13)	6.08 (1.89)	5.28 (1.73)	5.24 (1.83)	5.57 (2.11)	5.75 (2.26)	5.74 (2.29)	5.74 (2.32)	5.83 (2.41)	5.62 (2.63)
COMS	9.47 (2.05)	6.35 (0.91)	4.54 (0.62)	3.68 (0.58)	2.84 (0.95)	2.79 (1.01)	2.75 (1.12)	2.79 (1.21)	2.85 (1.29)	2.76 (1.33)
CPWR	9.12 (2.98)	6.98 (1.84)	5.62 (1.25)	4.87 (1.03)	4.10 (0.93)	3.91 (1.04)	3.86 (1.18)	3.86 (1.25)	3.87 (1.41)	3.77 (1.56)
CSCO	4.74 (0.75)	2.39 (0.31)	1.93 (0.39)	1.94 (0.42)	2.16 (0.52)	2.25 (0.57)	2.28 (0.56)	2.30 (0.58)	2.31 (0.61)	2.24 (0.65)
DELL	8.37 (1.93)	3.36 (0.82)	2.36 (0.68)	2.28 (0.74)	2.57 (0.84)	2.75 (0.93)	2.76 (0.82)	2.82 (0.83)	2.88 (0.88)	2.86 (1.02)
INTC	5.39 (1.28)	2.54 (0.54)	1.93 (0.39)	1.88 (0.44)	2.08 (0.51)	2.23 (0.55)	2.30 (0.57)	2.33 (0.61)	2.36 (0.63)	2.33 (0.68)
MSFT	4.31 (0.86)	1.97 (0.30)	1.60 (0.28)	1.61 (0.34)	1.85 (0.42)	1.94 (0.45)	1.95 (0.48)	1.98 (0.49)	2.01 (0.51)	2.03 (0.61)
NOVL	8.08 (1.60)	6.38 (0.91)	5.20 (0.75)	4.49 (0.72)	3.50 (0.68)	3.15 (0.60)	3.03 (0.67)	2.96 (0.77)	2.92 (0.84)	2.86 (0.86)
NXTL	6.66 (1.77)	5.07 (1.10)	4.10 (0.87)	3.55 (0.81)	3.01 (0.79)	2.81 (0.84)	2.72 (0.92)	2.68 (1.00)	2.72 (1.10)	2.65 (1.17)
ORCL	9.61 (2.50)	5.22 (1.12)	3.49 (0.66)	2.83 (0.71)	2.69 (0.85)	2.91 (0.91)	3.00 (0.95)	3.01 (1.03)	3.06 (1.06)	2.89 (1.14)
PSFT	11.70 (2.72)	9.51 (1.81)	7.81 (1.27)	6.77 (1.12)	5.06 (1.05)	4.33 (1.03)	3.93 (1.15)	3.73 (1.31)	3.60 (1.47)	3.40 (1.44)
QWST	7.59 (2.08)	5.12 (0.99)	3.87 (0.78)	3.39 (0.88)	3.07 (1.02)	3.13 (1.18)	3.20 (1.22)	3.27 (1.29)	3.31 (1.37)	3.19 (1.46)
SBUX	6.77 (2.01)	5.60 (1.39)	4.85 (1.13)	4.38 (1.01)	3.65 (0.92)	3.33 (0.86)	3.12 (0.89)	3.01 (0.98)	2.94 (0.94)	2.79 (1.13)
SUNW	5.56 (1.33)	3.37 (0.56)	2.63 (0.45)	2.51 (0.49)	2.68 (0.65)	2.76 (0.69)	2.84 (0.79)	2.85 (0.80)	2.83 (0.85)	2.80 (1.00)
WCOM	4.32 (0.66)	2.65 (0.44)	1.90 (0.42)	1.72 (0.40)	1.73 (0.46)	1.83 (0.51)	1.90 (0.57)	1.93 (0.55)	1.96 (0.57)	1.87 (0.66)

Note: This Table reports average daily realized volatilities ( $\overline{r}_{C,q}^{(1 \text{ sec})}$ ) sampled at a 1 second sampling frequency, corrected for different lags lengths ( $q$ ).

$$\overline{r}_{C,q}^{(1 \text{ sec})} = \frac{1}{N} \sum_{i=1}^N \sqrt{\sum_{s=1}^{h-1} \{r_{C,t-1+ih}^{(1 \text{ sec})}\}^2 + 2 \sum_{s=1}^q \frac{h-1}{h-1-s} \sum_{i=1}^{h-1-s} r_{C,t-1+ih}^{(1 \text{ sec})} r_{C,t-1+(i+s)h}^{(1 \text{ sec})}}$$

where  $N = 124$  are the total number of trading days. Standard deviations of  $\overline{r}_{C,q}^{(1 \text{ sec})}$  over the 124 trading days are reported in brackets.

Figure 5.1: Autocorrelation Functions for observed returns sampled at 1 second



Note: These graphs show the autocorrelation function for four representative stocks in the sample, sampled at 1 second intervals.

minute  $rv$ . This may again represent the trade-off point between the bias that is induced by the microstructure noise and the measurement error that is made in  $rv$ . However, when corrections are done over lags longer than the one that matches with 30 minute  $rv$ , we often observe that  $rv$  drops below the 30 minute  $rv$ , after which it corrects upwards again to a number close to the 30 minute level. A clear example of this is Dell, which drops below the 30 minute level of  $rv$  after a 10 second correction and moves back to the 30 minute  $rv$  reaching it again after having applied a 60 to 90 second correction. Even more interesting is the fact that the lowest value for  $rv$  found at a 15 second correction is close to the  $rv$  found with the structural model. These effects are more or less present for all the stocks in the sample but are more pronounced for the liquid stocks than the illiquid ones.<sup>8</sup>

The impact of the correction on  $rv$  is best observed when considering the autocorrelation function for the observed 1 second returns. In figure 5.1 we plot the autocorrelation function up to 50 lags for four representative stocks in the sample. All stocks display the negative autocorrelation induced by the microstructure noise. For the less liquid stocks like Apple and Amgen, we observe that this negative autocorrelation persists longer than for the more liquid stocks (Amazon and Dell). At longer lags the two liquid stocks reveal some positive autocorrelations, which can explain the upward movement observed in  $rv$ .

Next we also apply the correction in tick time. In table 5.8 we present the results for average  $rv$  when sampling at every transaction and correcting for different lag lengths. Standard deviations of average  $rv$  are reported in brackets.

The results for the correction in tick time reveal a very different pattern from the results obtained with calendar time sampling. For all of the stocks we find the lowest value for  $rv$  at a one or five tick correction on the data. Moreover, these values for  $rv$  closely correspond to the values for  $rv$  found in the structural model (table 5.6). Finding  $rv$  that closest approximates the 30 minute realized volatility leads in most cases to a minimum correction of 20 to 30 ticks. Correcting beyond 30 ticks does not have an impact on the level of  $rv$ . Hence, there seems to be persistence in tick returns of up to about 30 transactions. Another remarkable feature is that the standard deviations for average  $rv$  are almost always lowest when  $rv$  is lowest.

These results indicate that there are two sources affecting returns in tick time. The first source is the microstructure noise, which induces negative autocorrelations at very short lag lengths and causes an upward bias in uncorrected  $rv$ . The second one is a momentum in returns, which leads to positive autocorrelation up to much longer lags. This decreases the variance of  $rv$  when the correction is done over short lags lengths. These two effects clearly have different lag lengths at which they manifest themselves.

In figure 5.2 we plot the autocorrelation function for the four stock that we considered

---

<sup>8</sup>Likely, the correction for the illiquid stock needs to be done over even longer lags to observe this pattern.

Table 5.8: Realized Volatility: Tick time corrected

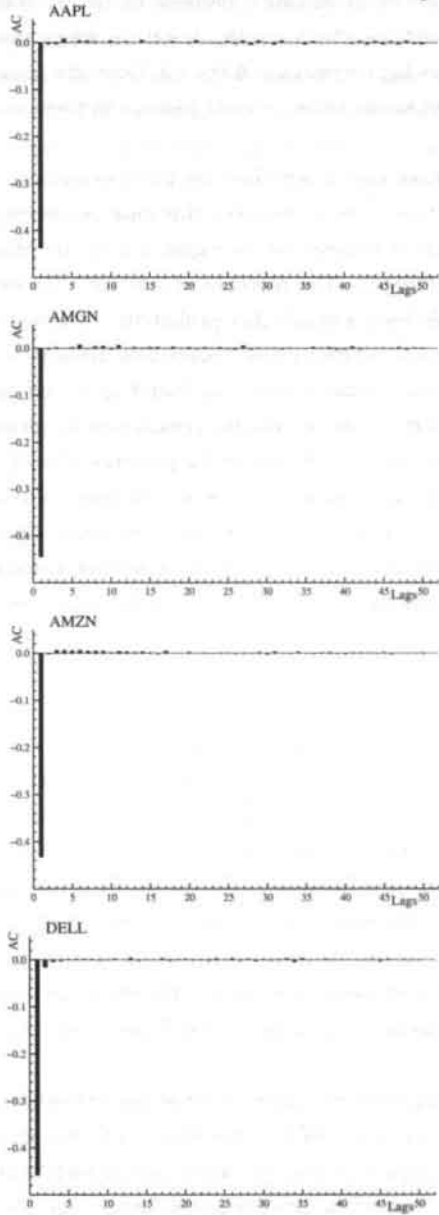
	q = 1	q = 5	q = 10	q = 20	q = 30	q = 40	q = 50	q = 100
AAPL	2.17 (0.56)	2.21 (0.54)	2.42 (0.66)	2.58 (0.71)	2.68 (0.75)	2.66 (0.79)	2.61 (0.78)	2.67 (0.94)
AMAT	2.35 (0.47)	2.61 (0.56)	2.96 (0.67)	3.23 (0.76)	3.36 (0.84)	3.35 (0.87)	3.34 (0.89)	3.37 (0.97)
AMGN	1.98 (0.55)	2.00 (0.56)	2.22 (0.62)	2.50 (0.77)	2.55 (0.83)	2.56 (0.84)	2.61 (0.88)	2.59 (0.97)
AMZN	3.56 (0.99)	4.01 (0.94)	4.48 (1.15)	4.84 (1.32)	4.85 (1.35)	4.89 (1.38)	4.95 (1.45)	4.94 (1.55)
ATHM	3.68 (1.10)	3.89 (1.22)	4.35 (1.43)	4.73 (1.65)	4.72 (1.81)	4.75 (1.82)	4.76 (1.87)	4.74 (1.96)
CMGI	4.37 (1.67)	4.76 (1.75)	5.36 (2.07)	5.68 (2.27)	5.74 (2.32)	5.71 (2.38)	5.71 (2.42)	5.70 (2.58)
COMS	2.79 (0.81)	2.31 (0.83)	2.44 (1.04)	2.65 (1.16)	2.84 (1.29)	2.85 (1.33)	2.76 (1.34)	2.74 (1.35)
CPWR	3.44 (1.25)	3.23 (1.10)	3.32 (1.14)	3.59 (1.35)	3.75 (1.44)	3.74 (1.46)	3.80 (1.47)	3.74 (1.64)
CSCO	1.71 (0.37)	1.81 (0.42)	2.04 (0.51)	2.21 (0.55)	2.26 (0.57)	2.30 (0.59)	2.30 (0.63)	2.25 (0.64)
DELL	2.69 (1.55)	2.15 (0.74)	2.32 (0.71)	2.67 (0.81)	2.78 (0.84)	2.82 (0.83)	2.80 (0.85)	2.76 (0.96)
INTC	1.72 (0.47)	1.74 (0.41)	1.96 (0.51)	2.20 (0.56)	2.33 (0.58)	2.34 (0.59)	2.34 (0.64)	2.33 (0.64)
MSFT	1.52 (0.40)	1.53 (0.35)	1.70 (0.36)	1.90 (0.47)	1.99 (0.48)	1.98 (0.50)	2.02 (0.50)	2.03 (0.58)
NOVL	2.78 (0.71)	2.40 (0.61)	2.58 (0.68)	2.74 (0.86)	2.83 (0.86)	2.87 (0.93)	2.94 (0.95)	2.91 (0.98)
NXTL	2.36 (0.90)	2.20 (0.72)	2.41 (0.91)	2.63 (1.06)	2.67 (1.11)	2.66 (1.17)	2.69 (1.11)	2.74 (1.21)
ORCL	2.75 (0.87)	2.37 (0.78)	2.47 (0.93)	2.89 (0.96)	2.92 (1.07)	2.92 (1.07)	3.02 (1.09)	2.85 (1.14)
PSFT	3.96 (1.33)	3.21 (1.18)	3.33 (1.27)	3.57 (1.49)	3.50 (1.45)	3.46 (1.54)	3.41 (1.62)	3.28 (1.47)
QWST	2.68 (0.86)	2.61 (0.87)	2.86 (1.10)	3.17 (1.21)	3.20 (1.25)	3.23 (1.30)	3.21 (1.35)	3.32 (1.41)
SBUX	2.62 (0.88)	2.54 (0.85)	2.51 (0.97)	2.73 (1.00)	2.86 (1.06)	2.85 (1.08)	2.90 (1.13)	2.94 (1.17)
SUNW	2.07 (0.53)	2.29 (0.54)	2.58 (0.64)	2.80 (0.72)	2.83 (0.80)	2.82 (0.85)	2.82 (0.90)	2.82 (0.96)
WCOM	1.52 (0.45)	1.51 (0.43)	1.68 (0.48)	1.87 (0.56)	1.92 (0.54)	1.91 (0.57)	1.89 (0.59)	1.87 (0.63)

Note: This Table reports average daily realized volatilities ( $\overline{RV}_{T,q}^{(1 \text{ tick})}$ ) sampled at every tick, corrected for different lags lengths ( $q$ ),

$$\overline{RV}_{T,q}^{(1 \text{ tick})} = \frac{1}{N} \sum_{t=1}^N \sqrt{\sum_{j=1}^{k(t)-1} \{r_{T,t-1+jk(t)}^{(1 \text{ tick})}\}^2 + 2 \sum_{s=1}^q \frac{k(t)-1}{k(t)-1-s} \sum_{j=1}^{k(t)-1-s} r_{T,t-1+jk(t)}^{(1 \text{ tick})} r_{T,t-1+(j+s)k(t)}^{(1 \text{ tick})}},$$

where  $N = 124$  are the total number of trading days. Standard deviations of  $\overline{RV}_{T,q}^{(1 \text{ tick})}$  over the 124 trading days are reported in brackets.

Figure 5.2: Autocorrelation Functions for observed returns sampled at every transaction



*Note:* These graphs show the autocorrelation function for four representative stocks in the sample, sampled at every transaction.

before. The autocorrelation functions confirm the findings in table 5.8. At the first lag we observe a large negative autocorrelation induced by the microstructure noise. There are several lags beyond the first lag which are also negative, which can explain the lowest value of  $rv$  to be found at a five lag correction. Although there are some positive autocorrelation found at longer lags, there seems to be no clear pattern in these autocorrelations (except for Amazon).

The question that arises now is whether the positive autocorrelation that decreases  $rv$  is a source of microstructure noise or whether this long persistency is inherent in the price process. Some persistence in returns can be explained by the presence of informed traders (see Glosten and Harris (1988)). The presence of informed traders causes order flow to be positively correlated. This results in a higher probability of a buy transaction to be followed by another buy transaction, which causes transaction returns to display autocorrelation. However, this positive autocorrelation was only found up to five lags in transaction returns (see Hasbrouck and Ho (1987)). As we find the persistence to pertain up to the thirtieth lag, it is doubtful whether this can be explained by the presence of informed traders. Moreover, as the autocorrelations persist up to such long lags, it is unlikely to be a source of microstructure noise. However, if this persistence is no microstructure noise, then the martingale property assumed for the price process does not hold. It is further remarkable that many of these small positive, unstructured autocorrelations have such a huge impact on the level of  $rv$ .

## 5.6 Conclusion

In this chapter we analyze Realized Variance in the presence of non- i.i.d. microstructure noise. Most research relies on the assumption that the microstructure noise is i.i.d., but this is shown not to hold in practice. In this chapter we follow two approaches to decompose the microstructure noise from the price process.

Firstly, we take a structural approach that allows for potential correlation between the microstructure noise and the innovation in the efficient price. Using ultra-high frequency transaction data, we decompose the stock price into a component referring to the efficient price and one referring to microstructure noise. However, the realized volatilities found at the highest sampling frequencies are substantially lower than the realized volatility at lower frequencies.

We follow a second approach to correct for the microstructure noise, that was recently introduced by Hansen and Lunde (2004b). We show that when sampling is done in calendar time, we mostly observe negative autocorrelation, which lead to an upward biased measure for the realized volatility, even when correcting over longer lags. We show that these negative autocorrelations are induced by the way that we sample.

Applying the correction in tick time shows that most of the negative autocorrelation is

centered at the first or the second lag. Applying this approach with a one lag correction leads to similar results as found in the structural approach. We further find that correction for longer lags in tick time increases the realized volatility up to a level equal to the low frequency realized volatilities. As the lag length needed for realized volatility to be equal to the low frequency realized volatility (up to 30 lags), we wonder whether this is microstructure noise or whether this is an inherent persistency in the price process.



## 5.A Appendix: Identification of the univariate state space model

This appendix addresses the identification of the state space model (5.17). We start with the equation for log prices

$$\begin{aligned}
 p_{t+jk(t)} &= p_{t+jk(t)}^* + u_{t+jk(t)}, \\
 p_{t+jk(t)}^* &= p_{t+(j-1)k(t)}^* + \tau_{t+jk(t)}^{\delta_1} \varepsilon_{t+jk(t)}, \\
 u_{t+jk(t)} &= \alpha \tau_{t+jk(t)}^{\delta_2} \varepsilon_{t+jk(t)} + D_{1,t+jk(t)} e_{1,t+jk(t)} + \\
 &D_{2,t+jk(t)} e_{2,t+jk(t)} + D_{3,t+jk(t)} e_{3,t+jk(t)}.
 \end{aligned} \tag{5.22}$$

The three idiosyncratic noise terms  $e_{1,t+jk(t)}$ ,  $e_{2,t+jk(t)}$  and  $e_{3,t+jk(t)}$  depend on the time of the day. The dummy variable  $D_{1,t+jk(t)}$  equals 1 between 9.30 and 11.00 and otherwise zero,  $D_{2,t+jk(t)}$  equals 1 between 11.00 and 14.30 and otherwise zero and  $D_{3,t+jk(t)}$  equals 1 between 14.30 and 16.00 and otherwise zero.

On day  $t$ , consider we are between 9.30 and 11.00. Returns are then given as

$$\begin{aligned}
 r_{t+jk(t)} &= p_{t+jk(t)} - p_{t+(j-1)k(t)} \\
 &= \tau_{t+jk(t)}^{\delta_1} \varepsilon_{t+jk(t)} + \alpha \tau_{t+jk(t)}^{\delta_2} \varepsilon_{t+jk(t)} + e_{1,t+jk(t)} \\
 &\quad - \alpha \tau_{t+(j-1)k(t)}^{\delta_2} \varepsilon_{t+(j-1)k(t)} - e_{1,t+(j-1)k(t)}.
 \end{aligned} \tag{5.23}$$

The variance of returns is given by

$$\begin{aligned}
 E[\{r_{t+jk(t)}\}^2] &= \\
 &\sigma^2(\tau_{t+jk(t)}^{2\delta_1} + 2\alpha\tau_{t+jk(t)}^{(\delta_1+\delta_2)} + \alpha^2(\tau_{t+jk(t)}^{2\delta_2} + \tau_{t+(j-1)k(t)}^{2\delta_2})) + 2\omega_1^2,
 \end{aligned} \tag{5.24}$$

where  $E[\varepsilon_{t+jk(t)}] = \sigma^2$  and  $E[e_{1,t+jk(t)}] = \omega_1^2$ . Consequently, first order autocorrelations are given as

$$\begin{aligned}
 E[\{r_{t+jk(t)}\}\{r_{t+(j-1)k(t)}\}] &= \\
 &-\sigma^2(\alpha\tau_{t+(j-1)k(t)}^{(\delta_1+\delta_2)} + \alpha^2\tau_{t+(j-1)k(t)}^{2\delta_2}) - \omega_1^2.
 \end{aligned} \tag{5.25}$$

To show that  $\alpha$  and  $\omega_1^2$  are uniquely identified consider a linear translation on  $\alpha$

$$\dot{\alpha} = \alpha + x. \tag{5.26}$$

Given this translation, the variance of returns is given by

$$\begin{aligned}
 E[\{r_{t+jk(t)}\}^2] &= \\
 &\sigma^2(\tau_{t+jk(t)}^{2\delta_1} + 2\dot{\alpha}\tau_{t+jk(t)}^{(\delta_1+\delta_2)} + \dot{\alpha}^2(\tau_{t+jk(t)}^{2\delta_2} + \tau_{t+(j-1)k(t)}^{2\delta_2})) + 2\dot{\omega}_1^2 \\
 &= \sigma^2(\tau_{t+jk(t)}^{2\delta_1} + 2\dot{\alpha}\tau_{t+jk(t)}^{(\delta_1+\delta_2)} + \dot{\alpha}^2(\tau_{t+jk(t)}^{2\delta_2} + \tau_{t+(j-1)k(t)}^{2\delta_2})) \\
 &\quad + \sigma^2(2x\tau_{t+jk(t)}^{(\delta_1+\delta_2)} + 2\dot{\alpha}x(\tau_{t+jk(t)}^{2\delta_2} + \tau_{t+(j-1)k(t)}^{2\delta_2}) + x^2(\tau_{t+jk(t)}^{2\delta_2} + \tau_{t+(j-1)k(t)}^{2\delta_2})) \\
 &\quad + 2\dot{\omega}_1^2.
 \end{aligned} \tag{5.27}$$

The first order autocorrelation is given by

$$\begin{aligned}
 E[\{r_{t+jk(t)}\}\{r_{t+(j-1)k(t)}\}] &= \\
 &-\sigma^2(\dot{\alpha}\tau_{t+(j-1)k(t)}^{(\delta_1+\delta_2)} + \dot{\alpha}^2\tau_{t+(j-1)k(t)}^{2\delta_2}) - \dot{\omega}_1^2 \\
 &= -\sigma^2(\dot{\alpha}\tau_{t+(j-1)k(t)}^{(\delta_1+\delta_2)} + \dot{\alpha}^2\tau_{t+(j-1)k(t)}^{2\delta_2}) \\
 &\quad - \sigma^2(x\tau_{t+(j-1)k(t)}^{(\delta_1+\delta_2)} + 2\dot{\alpha}x\tau_{t+(j-1)k(t)}^{2\delta_2} + x^2\tau_{t+(j-1)k(t)}^{2\delta_2}) - \dot{\omega}_1^2.
 \end{aligned} \tag{5.28}$$

As there is no linear translation found on  $\omega_1^2$  which solves for (5.27) and (5.28), the parameters  $\alpha$  and  $\omega_1^2$  are uniquely identified, as long as  $\delta_2 \neq 0$  and  $(\delta_1 + \delta_2) \neq 0$ . A further necessity is that  $Var(\tau_{t+jk(t)}) \neq 0$ .



## Chapter 6

# The Information Content of Trades and Quotes

### 6.1 Introduction

The theoretical literature on the presence of asymmetric information in financial markets is abundant (see e.g. Kyle (1985), Glosten and Milgrom (1985), and Glosten and Harris (1988), among others). With the general notion that prices change with the flow of information, a central question is where the information comes from. The theoretical literature recognizes two different sources of information. One source relates to market wide information about the performance of a company (public information) and one source refers to private information shared among few traders called insiders. A relevant question is to what extent these two sources of information affect the change in prices.

Most of the past empirical research has either focussed on the additional costs in prices that private information causes (Glosten and Harris (1988)), the probability of trading with informed traders (Easley, Kiefer, O'Hara, and Paperman (1996)) or the contributions to the price discovery process (Hasbrouck (1995)) of differently informed venues. Hence most interest has been drawn towards explaining the asymmetry and not so much the absolute magnitudes. Madhavan, Richardson, and Roomans (1997) do address public and private information separately, obtaining the private information component from the autocorrelation in order flows (see e.g. Glosten and Harris (1988)) and public information (corrected for some additional microstructure noise features) from the remaining variance in price changes.

The approach of Madhavan, Richardson, and Roomans (1997) is based on the following. Dealers set quotes at which traders can execute orders. If these prices reflect all available information (i.e. there is no private information) trades will be executed at both sides of the market with equal probability. If, however, there is an asymmetry in information, there is a higher probability that trades are executed on one side of the market. This generates

autocorrelation in the order flow. The autocorrelation pattern in order flows should therefore reveal the amount of private information.

This approach, however, relies heavily on a specific trade classification rule, such as the rule proposed by Lee and Ready (1991). Where this algorithm, to a large extent, correctly matches these trades for the NYSE, results for the Nasdaq, the market where this study focusses on, are poor (see Blume and Goldstein (1997)). Additionally, Madhavan, Richardson, and Roomans (1997) use only the quotes that precede trades. So quote innovations without trade occurring in between are not considered in their sample. This could lead to an upward bias in the fraction of private information measured using their methodology.

In this chapter we opt for a different approach, which fits the framework of Hasbrouck (1991). Instead of relying on a specific classification algorithm, *all* trades and *all* quotes are used to evaluate the amount of public and private information in prices. We rely on the fact that both trades and quotes share the same common trend, referred to as the *true* price. Whenever a trade or a quote innovation occurs, information about the location of the true price is released. Measuring the extent to which quotes or trades move the efficient price provides us with a measure for the amount of public and private information. Assuming that dealers are mere liquidity providers, their quotes will be based on the public information available.<sup>1</sup> Traders can react to these prices either for liquidity motives or based on private information they might have.

When a quote is innovated this is done based on the public information available. As the dealer, who innovates her quotes, sees the transactions prior to her quote innovation, her quote includes all the private information included in these transactions. However, as this private information has already been incorporated into the price process, the information coming from a quote update purely reflects the innovation in public information. The public information can thus be determined from the extent to which the quote innovations affect the true price process.

The amount of private information is obtained from the trade process. When quotes are set by dealers, traders can respond to these quotes by trading. If these quotes are set correctly, trades occur at both sides of the market with equal probability. In this case the contribution of trades to the innovation in the true price is zero. However, if there is private information among traders, trades are more likely to occur on one particular side of the market, revealing information for the direction of the true price process. Thus the extent to which the trade process affects the true price process reveals the amount of private information. This methodology is consistent with the idea that informed trading causes

<sup>1</sup>To the extent that Nasdaq dealers do quote strategically, we note that this information becomes public as soon as the dealer disseminates the quote to the market. Some evidence on the strategic nature of Nasdaq market makers has been proposed by Ellis, Michaely, and O'Hara (2002), Schultz (2003) and Chung and Cho (2004). However, this strategic behavior is directed mainly towards obtaining preferencing agreements with potential traders, instead of taking strategic positions in the market.

autocorrelation in order flow (Glosten and Harris (1988)), however, we do not rely on a specific classification algorithm.

We empirically evaluate the private and public information components of 20 actively traded companies listed at the Nasdaq over the period February 1<sup>st</sup> 1999 until July 31<sup>st</sup> 1999, covering a total of 124 trading days. As we propose a model constructed for data sampled at every observation (all trades and quotes that occur) we can uniquely identify the impact of both types of information. Overall we find that the private information component is very small in comparison to the public information component (causing on average only 10% of the daily price changes which confirms the findings of Madhavan, Richardson, and Roomans (1997)).

Since the data is sampled at every observation, the impact of time between observations on the change in the true price process can be measured. Overall we find that time has a small impact on the innovation in this price process, rejecting the notion that the price process evolves in calendar time and confirming the findings of Clark (1973) who promotes the evolution of the price process in a different time scale.

The data set used allows us to test for the impact of some additional factors. Firstly, we test whether high traded volumes are more informative than low traded volumes. We find some evidence that different trading volumes affect the amount of private information. However, quoted volume and especially the amount of dealers that set the inside (dealers that jointly set the best price) are much more informative. More dealers at the inside reveals more public information and hence less asymmetry. Finally, the fact whether traditional market makers or an Electronic Communication Networks (ECNs) set the inside matters. We find that when market makers are at the inside, quotes are more informative (less asymmetry) than when ECNs are at the inside.

Other factors that determine the price setting of dealers were investigated by Engle and Patton (2004). They observe that aggregate traded volume previous to the innovation in quotes has a permanent impact on the midpoint of the quotes. Their model can be seen as a model that describes the flow of private information into public information (eventually all private information will become public).

The remainder of this chapter proceeds as follows. The next section introduces a model for the trade and quote process, where both are modelled as they arrive in the market. Section 6.3 addresses the data used in this study. In section 6.4 results of the model are presented. Section 6.5 tests for the impact of additional factors in the model. Finally section 6.6 concludes.

## 6.2 Model

This section develops a model for the dynamics of trades and inside quotes (inside quotes are defined as the best bid and ask price submitted by all dealers). The model is designed for ultra-high frequency data, where we sample at every transaction or quote innovation that occurs within the market. The model we propose is an unobserved components model as in Hasbrouck (1993), where we decompose transaction prices and quote innovations into an efficient price component and a component related to transitory effects/microstructure noise. We first address transaction prices and subsequently consider the process for quotes. After this we integrate both processes into one single model.

On a given day there are  $N$  trades executed in a single stock, where  $n = 1, \dots, N$ .<sup>2</sup> Let  $p_n$  be the log price of the  $n^{\text{th}}$  transaction reported to the market at time  $t_n$ . This price can represent either a buy or a sell transaction. We decompose this transaction price into two components. The first part considers the efficient price from the transaction ( $m_n$ ). The second part considers the microstructure noise ( $v_n$ ) present in the data, such as the bid-ask bounce and other effects. The decomposition follows as

$$p_n = m_n + v_n, \quad (6.1)$$

similar to Hasbrouck (1993), Madhavan (2000) or Zhang, Mykland, and Aft-Sahalia (2003) among others.

On the same day that we observe  $N$  trades, we observe  $M$  quote innovations. As these quote innovations do not occur at the same point in time as the transactions, we index all quote innovation with  $m$ , where  $m = 1, \dots, M$ . The vector  $q_m$  contains all log quote prices that are innovated at time  $t_m$ . This vector can be one dimensional or two dimensional depending on whether bid, ask, or both quotes are innovated at time  $t_m$ . For the moment we assume that both quotes are innovated at  $t_m$  such that the vector of quote innovations has dimension 2. Later we relax this assumption to obtain the full model.

Similar to trades, we decompose quotes. As bid and ask quotes are characterized by a deterministic difference between bid and ask prices, due to e.g. order processing costs, we decompose this process into three components. The first component considers this deterministic part ( $c$ ), due to the fixed components in the bid-ask spread. The second component considers the efficient price, shared by both bid and ask quote ( $m_m$ ). The last component ( $\nu_m$ ) considers deviations from the efficient price that are not captured by  $c$ . These deviations from the efficient price can exist due to asymmetric information, inventory imbalances

---

<sup>2</sup>In fact  $N$  varies across trading days. However, we do not index this variable for the sake of notational clarity.

or other effects in the inside bid and ask. The decomposition for quotes reads

$$q_m = c + \iota m_m + \nu_m, \quad (6.2)$$

where  $\iota$  is a unit vector of dimension 2.

As trades and quotes are considered for the same asset, both processes are cointegrated and share the same common trend, the efficient price. We refer to the occurrence of either trade, quote or both as an *event*. Whenever one of these events occurs, information about the location of the efficient price is released. Therefore, this is the appropriate time scale for the evolution of the efficient price and we refer to this time scale as *event time*. Let  $t_\ell$  be the time that we observe any of these events, where  $t_\ell = t_n \cup t_m$  and  $\ell = 1, \dots, L$ . The efficient price, common to both trades and quotes evolves in this time scale and is assumed to follow a random walk,

$$m_\ell = m_{\ell-1} + \sigma_\ell \varepsilon_\ell, \quad (6.3)$$

where  $\varepsilon_\ell$  is an i.i.d. innovation term with unit variance. The volatility of the random walk ( $\sigma_\ell$ ) is allowed to depend on several factors. The factor we are mainly interested in is the source of the innovation in the efficient price. This source can be an innovation in quotes or the occurrence of a trade. Additionally, we are interested in the impact of time itself on the innovation in the efficient price. We allow the size of the innovation in the random walk to depend on the duration between events ( $\tau_\ell$ ). The duration is defined as the time between events divided by the average duration (over the whole sample) between events. This normalization causes the volatility parameters to be interpretable per event, instead of per second. All other parameters are not affected by this normalization. The full specification for  $\sigma_\ell$  reads

$$\sigma_\ell = \tau_\ell^{\delta_1} (P_\ell(1 - Q_\ell)\sigma_P + Q_\ell(1 - P_\ell)\sigma_Q + P_\ell Q_\ell \sigma_{PQ}), \quad (6.4)$$

where  $\delta_1$  measures the impact of time between events on the innovation in the random walk. The variable  $P_\ell$  is an indicator variable equal to 1 when a trade has occurred at time  $t_\ell$  and zero otherwise. Likewise,  $Q_\ell$  is equal to 1 when at time  $t_\ell$  a quote has been innovated and zero otherwise. The impact of a transaction on the innovation in the random walk is measured by  $\sigma_P$ , the innovation due to a release of a new quote is measured by  $\sigma_Q$ , and when both events occur we capture the innovation in  $\sigma_{PQ}$ .<sup>3</sup>

Traditionally, dealers are assumed to have access to the public information and derive some additional information from the net order flow they receive from trades. Dealers therefore operate solely as order processors who set clearing prices, based on public information, that maintain a stable inventory position. In this respect we expect the innovation in the effi-

<sup>3</sup>We later on allow for more dynamics in these parameters, by testing the impact of several factors on these parameters.



cient price due to quotes to be equal to the innovation in market wide beliefs. However, recent research (Ellis, Michaely, and O'Hara (2002), Schultz (2003) and Chung and Cho (2004)) has focussed on the strategic nature of traditional market makers. They have shown that market makers engaged in the IPO process, or work for a company with analysts investigating the particular stock have a clear dominance in this stock over the other market makers. Most of these strategic activities, however, are directed towards obtaining preferencing agreements and not so much towards obtaining private information. On the other hand, some recent research has indicated that there exist informational asymmetries among Nasdaq dealers (see e.g. Huang (2002), and Frijns and Schotman (2004)). To this extent we note that informational asymmetry may be present among dealers, but this asymmetry disappears when the dealer issues her quote. We therefore argue that the innovation in the efficient price due to quotes, reflects the public information in the market.

Traditional microstructure theory on asymmetric information (e.g. Glosten and Milgrom (1985), Easley and O'Hara (1987), and Easley and O'Hara (1992)) considers two different investors in the market, informed traders and noise traders. Noise traders are often assumed to trade for liquidity purposes and have no additional information than the information in quoted prices, the public information. Informed traders, derive information from additional sources and trade when the true price of an asset lies above or below quoted prices. As only private information leads to strategic trading (causing an autocorrelation in order flow), this will have an impact on the innovation in the efficient price ( $\sigma_P$ ).<sup>4</sup>

Another factor considered in (6.4) is the impact of time on the innovation in the efficient price. Clark (1973) was the first to point out that physical time might not be the appropriate time scale for the price processes to evolve in. He shows that the distribution of price changes can better be described as a subordinate to the normal distribution, with traded volume as the directing process. More recently Ané and Geman (2000) have shown that it is not so much traded volume that directs the distribution of price changes, but that the number of transactions in an interval is a more appropriate time scale, a notion already proposed by Harris (1987).

We measure the impact of time on the innovation in the random walk by  $\delta_1$ . When  $\delta_1 = \frac{1}{2}$  the variance of the random walk is proportional to the duration between trades or quotes. In this case the process is said to evolve in physical time or so called calendar time. When  $\delta_1 = 0$  the random walk evolves in event time. In this case the innovation in the random walk then does not depend on the time between events but only on the fact that an event has taken place. For the innovation in the random walk we consider the duration between events as the relevant time scale, as at each event information is released.

<sup>4</sup>One could argue that the impounding of trades on one side of the market, leading to an increase or a decrease in the inventory of the respective market maker(s), ultimately leads to the adjustment of dealer quotes, making the private information public.

Additionally we want to specify the microstructure noise in trades and quotes. As the size of the microstructure noise might be informative about the size of the innovation in the efficient price, we allow both noise terms to correlate with the innovation in the efficient price. These correlations are also allowed to depend on time. We include duration functions to measure the impact of time on the correlations. Here we take, instead of the time between events, the time between trades or quotes, for trades and quotes respectively. Although the impact of time between events will also be interesting, we prefer to look at quote durations or trade durations separately. We start again by considering trades. The durations between trades ( $\pi_{P,\ell}$ ) are defined similar to durations between events. They are defined as the time between transactions divided by the average duration of transactions. We decompose the microstructure noise for trades as

$$v_\ell = \alpha \pi_{P,\ell}^{\delta_2} (P_\ell(1 - Q_\ell)\sigma_P + P_\ell Q_\ell \sigma_{PQ}) \varepsilon_\ell + u_\ell. \quad (6.5)$$

The parameter  $\alpha$  measures the covariance between the innovation in the efficient price and the noise in transactions. The noise in this case can only correlate with the innovation due to trades  $\sigma_P$  or with the innovation due to both events  $\sigma_{PQ}$ . To measure the time impact on this correlation we consider the time between consecutive transactions, instead of the time between events. For this purpose this is the relevant time scale as transactions are only observed at these points in time. The parameter  $\delta_2$  measures the impact that time has on this correlation. When  $\delta_2 > 0$  long durations between transactions lead to a higher correlation between the innovation in the efficient price and the noise and vice versa. The last part in the equation ( $u_\ell$ ) is i.i.d. noise.

Likewise we specify the temporary deviations from the efficient price for the inside quotes. Under the assumption that both quotes are observed at  $t_\ell$  we have

$$v_\ell = \beta \pi_{Q,\ell}^{\delta_3} (Q_\ell(1 - P_\ell)\sigma_Q + P_\ell Q_\ell \sigma_{PQ}) \varepsilon_\ell + e_\ell, \quad (6.6)$$

where  $\beta$ , a 2-vector, measures the correlation between the quote noise and the innovation in the efficient price. Similar to the noise of the trade process, the quote noise only correlates with the innovation in the efficient price due the trades  $\sigma_Q$  or the innovation due to both events  $\sigma_{PQ}$ . The relevant durations for quotes are the times between quote innovations ( $\pi_{Q,\ell}$ ). These durations are again normalized by the average duration between quotes. As before the impact time has on this correlation is measured by  $\delta_3$ . The last part in the equation ( $e_\ell$ ) measures idiosyncratic noise.

In the discussion above we have assumed that all events are observed at every point in time  $t_\ell$ . However, as mentioned, both events do not always occur at the same point in time. To solve this issue, let  $y_\ell$  be a  $3 \times 1$  observation vector that includes the most recent prices and

quotes at time  $t_\ell$ . If at time  $t_\ell$  a transaction is observed this vector includes this transaction and the standing quotes at  $t_\ell$ . Likewise when a new quote is released this is included in  $y_\ell$  as well as the other standing quote and the last observed transaction. As we only want to consider the new information available we define the matrix  $J_\ell$  as the matrix that selects the observations from  $y_\ell$  that are new at  $t_\ell$ . This matrix has dimensions  $(s_\ell \times 3)$  where  $s_\ell$  is the number of new observations at  $t_\ell$ .

As an illustration, consider a transaction that occurs at the same time as an innovation in the bid quote. This trade and quote are selected from the observation vector as follows,

$$J'_\ell y_\ell = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_\ell \\ ask_\ell \\ bid_\ell \end{pmatrix} = \begin{pmatrix} p_\ell \\ bid_\ell \end{pmatrix}.$$

This pre-multiplication with the selection matrix can be applied to all the components in the model.

With this pre-multiplication we can now write the model in a state space framework. We obtain

$$\begin{aligned} J_\ell y_\ell &= J_\ell \begin{pmatrix} 0 \\ c \end{pmatrix} + J_\ell \iota m_\ell + J_\ell \begin{pmatrix} v_\ell \\ \nu_\ell \end{pmatrix}, \\ m_\ell &= m_{\ell-1} + \sigma_\ell \varepsilon_\ell, \\ J_\ell \begin{pmatrix} v_\ell \\ \nu_\ell \end{pmatrix} &= J_\ell \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \odot \begin{pmatrix} 1 & 0 \\ 0 & \iota \end{pmatrix} \begin{pmatrix} \pi_{P,\ell}^{\delta_2} (P_\ell(1-Q_\ell)\sigma_P + P_\ell Q_\ell \sigma_{PQ}) \\ \pi_{Q,\ell}^{\delta_3} (Q_\ell(1-P_\ell)\sigma_Q + P_\ell Q_\ell \sigma_{PQ}) \end{pmatrix} \varepsilon_\ell + J_\ell \begin{pmatrix} u_\ell \\ e_\ell \end{pmatrix}, \end{aligned} \quad (6.7)$$

with  $\mathbf{0}$  and  $\iota$  vectors of dimension  $(2 \times 1)$ .

We can estimate all parameters using a Kalman Filter. We estimate the model by QML according to Harvey (1989), where we treat the observation that are not selected by  $J_\ell$  as missing observations. As the latent process in the model is a random walk we initialize the system with a diffuse prior. Although the parameters are estimated over the whole sample, at the beginning of each day the system is re-initialized, by setting the variance of the prediction error to a very large number. This entails that the first 50 observations of each day are not considered in the calculation of the likelihood function, but are used to let the prediction error variance decrease to its normal level.

Since the innovation in the latent process depends on the source of the information and on the duration between events, the model does not converge to a steady state.<sup>5</sup> The variance of the prediction error is different at each observation. Due to this all the calculations in the Kalman Filter recursions have to be done. This increases the computational efforts needed

<sup>5</sup>Note that the parameters in  $\alpha$  and  $\beta$  are only identified as we allow for these time variations.

Table 6.1: List of Company Ticker symbols and Company names

Ticker Symbol	Company name
AAPL	Apple Computer Inc.
AMAT	Applied Materials Inc.
AMGN	Amgen Inc.
AMTD	Ameritrade Holding Corporation
ATHM	At Home Corporation
CIEN	Ciena Corporation
COMS	3Com Corporation
CPWR	Compuware Corporation
CSCO	Cisco Systems Inc.
DELL	Dell Computer Corporation
INTC	Intel Corporation
MSFT	Microsoft Corporation
NOVL	Novell Inc.
NXTL	Nextel Communications CL-A
ORCL	Oracle Corporation
PSFT	Peoplesoft Inc.
QWST	Qwest Communications Intl Inc.
SBUX	Starbucks Corporation
SUNW	Sun Microsystems Inc.
WCOM	MCI WorldCom Inc.

for the estimation of the model significantly.<sup>6</sup>

### 6.3 Data

This section addresses the trade and quote data analyzed in the empirical part of this study and discusses some summary statistics of the data. In this study we consider 20 actively traded stocks at Nasdaq. The sample period extends from February 1<sup>st</sup> 1999 until July 31<sup>st</sup> 1999, spanning a total of 124 trading days. The stocks are selected randomly, but with varying liquidity and were all included in the Nasdaq-100 index at some stage during 1999. In table 6.1 we report the company names and their ticker symbols. These symbols will be used in the tables in the remainder of this chapter. Trade and quote data were provided by Nasdaq. We first address the trade data and subsequently address quotes.

The trade data provided by Nasdaq contains all transactions at the Nasdaq National Market within regular trading hours (9.30 - 16.00). This data set includes the time the trade was reported, the time the trade was executed, price and volume of the trade, and some indicators. All these indicators refer to late reported trades and these trades are removed from the data. The inclusion of the executed times in the data set provides a great benefit. It is known that high frequent intra-day stock prices are contaminated with microstructure

<sup>6</sup>As we select different observations at each point and are thus faced with different correlation between observations, treating the multivariate system as a univariate one, as proposed by Koopman and Durbin (2000) does not yield computational advantages.

noise. Issues such as the bid ask bounce, price discreteness and non-synchronous trading, contaminate data. One major factor contributing to the microstructure noise are the times that trades are reported. In many cases we observe that transactions executed at the same point in time are reported at different times to the system. Although we opt for reported time as the relevant time scale (as this is the point in time that the information is released to the market) many reported transactions are non-informative, as later executed trades have been reported already. Therefore, we remove all transactions that are reported to the system after a trade has been reported that was executed later. Likewise, when trades are executed at the same point in time, we only consider the transaction that was reported first. This removes a substantial amount of data and minimizes the discrepancy between executed time and reported time. For the first part of the analysis we only consider the price data. Later, when the model is extended we also include volumes.

The quotes considered in this study are the inside quotes in the market. The quotes are the best bid and offer quotes issued by all market makers and Electronic Communication Networks (ECNs). The inside quote data contains the time that the new quotes are issued at the market, the best bid and ask quotes, and some indicators referring to the status of the quote. From the quote data we remove all non-regular quotes (these are quotes with a so called Halt indicator). We consider all bid and ask quotes more than \$2 away from the average of the past 50 bid or ask quotes, respectively, as outliers and remove these quotes from the data set. We further remove all quotes that do not change. For example when the bid quote is innovated, this is the only observation that we consider and the other quote is removed. Sampling in this manner has clear advantages to considering the mid quote as a change in the mid quote does not identify whether this occurred due to a change in the bid or the ask quote.

In addition to the inside quotes also the individual dealer quotes are used. Characteristics of the dealer quote data are used for the extension of the model in a later part of this chapter. We will address the relevant data issues there as well.

Table 6.2 reports some summary statistics on the trade and quote data used in the empirical part of this chapter. The first column reports the total number of observations over the 124 trading days. These numbers indicate the large range of liquidity that we span. Where for the most liquid stock, Dell, we observe 1,359,451 observations, we only observe 248,247 for the least liquid stock, Starbucks. The next column reports the fraction of times that the only observation is a transaction. Clearly, there are many more transactions observed than quotes, with an average percentage of trades around 80%. This clearly contrasts observation from the NYSE, where more quotes are observed than trades. One explanation for this might be that the NYSE is an order driven market, i.e. the specialist sets quotes based on the orders that arrive. This results in the fact that for almost every transaction there is an innovation in quotes. Furthermore, the inside quotes reported at the Nasdaq do not include quoted volume.

Table 6.2: Summary statistics

	Number of Observations	Percentage of Trades	Percentage of Quotes	Observation Duration	Quote Duration	Trade Duration
AAPL	345,733	79.76 %	16.76%	8.37	61.66	11.08
AMAT	769,143	77.47 %	16.43%	3.77	24.89	4.76
AMGN	476,821	75.45 %	20.03%	6.08	32.92	7.87
AMTD	675,743	58.07 %	30.57%	4.29	12.71	6.93
ATHM	878,388	63.02 %	25.87%	3.30	11.08	4.76
CIEN	368,267	79.69 %	16.60%	7.86	57.15	10.58
COMS	572,440	92.04 %	5.98%	5.06	110.73	5.74
CPWR	358,903	81.79 %	15.24%	8.06	65.84	10.37
CSCO	1,178,831	81.46 %	11.44%	2.46	20.49	2.77
DELL	1,359,451	92.76 %	3.82%	2.13	54.42	2.22
INTC	1,166,695	87.03 %	7.94%	2.48	31.96	2.69
MSFT	1,274,909	85.51 %	8.54%	2.27	25.55	2.46
NOVL	322,087	88.86 %	9.55%	8.97	133.62	10.88
NXTL	376,470	83.96 %	13.07%	7.69	75.51	9.44
ORCL	841,023	92.69 %	5.16%	3.44	83.17	3.74
PSFT	278,968	90.34 %	8.29%	10.36	179.41	12.39
QWST	596,458	79.31 %	16.11%	4.86	32.76	6.15
SBUX	248,247	79.49 %	17.69%	11.66	80.68	16.00
SUNW	864,572	77.68 %	15.71%	3.35	21.78	4.09
WCOM	805,931	84.43 %	11.35%	3.60	35.55	4.08

*Note:* This table reports some summary statistics on the raw data and the filtered data. The first column reports the total number of filtered observations over the 124 trading days. The second column reports the fraction of times only a trade is observed. The third column reports the fraction of times only a quote is observed. The last three columns refer to average durations. The first of these columns represents the average duration between any observation. The next column reports the average quote duration and the last column reports the average trade duration.

The next column reports the fraction of quote innovations. These fractions are quotes over total observations, where no transactions are observed. We find, on average, that quotes represent only about 14% of the observations, leaving a remaining 6% of the observation where both a trade and a quote innovation was observed. Finally, the last three columns report the duration between observations, quotes and trades respectively. These durations again emphasize the difference in the liquidity of the assets we consider in this study, where the duration between quote innovation (on average approximately 60 seconds) is much larger than the duration between transactions (on average about 7 seconds).

## 6.4 Results

This section addresses the results of the model proposed in section 6.2. We address the total contribution that trades and quotes have in daily price changes, but start by discussing the parameter estimates of (6.7).

Table 6.3: Duration parameters and Correlations

Symbol	$\delta_1$	$\delta_2$	$\delta_3$	$\alpha$	$\beta_b$	$\beta_a$
AAPL	0.01 (0.01)	-0.02 (0.06)	-0.09 (0.04)	0.21 (0.02)	-0.05 (0.01)	-0.08 (0.01)
AMAT	-0.03 (0.01)	-0.08 (0.05)	0.21 (0.05)	0.26 (0.02)	-0.07 (0.01)	-0.10 (0.01)
AMGN	-0.04 (0.01)	-0.04 (0.06)	-0.40 (0.15)	0.27 (0.02)	-0.03 (0.01)	-0.03 (0.02)
AMTD	-0.13 (0.01)	-0.05 (0.04)	0.09 (0.11)	0.69 (0.05)	0.04 (0.02)	0.04 (0.03)
ATHM	-0.09 (0.01)	0.00 (0.03)	0.19 (0.06)	0.55 (0.03)	0.06 (0.01)	0.03 (0.02)
CIEN	-0.01 (0.01)	-0.04 (0.06)	-0.23 (0.07)	0.25 (0.02)	-0.01 (0.01)	-0.05 (0.01)
COMS	-0.06 (0.02)	-0.30 (0.55)	-0.66 (0.48)	0.07 (0.07)	-0.09 (0.17)	0.00 (0.01)
CPWR	-0.07 (0.01)	-0.15 (0.11)	-0.38 (0.10)	0.24 (0.05)	-0.03 (0.01)	-0.04 (0.02)
CSCO	-0.08 (0.01)	0.03 (0.03)	-0.38 (0.24)	0.35 (0.01)	0.00 (0.01)	-0.01 (0.01)
DELL	-0.07 (0.02)	-0.24 (0.28)	-0.48 (0.06)	0.12 (0.03)	-0.04 (0.01)	-0.04 (0.01)
INTC	-0.05 (0.02)	-0.26 (0.66)	-0.64 (0.29)	0.11 (0.08)	-0.03 (0.03)	-0.09 (0.08)
MSFT	-0.08 (0.07)	-0.25 (1.54)	-0.47 (0.06)	-0.02 (0.04)	-0.11 (0.03)	-0.38 (0.09)
NOVL	0.07 (0.01)	0.11 (0.13)	0.01 (0.16)	0.15 (0.03)	-0.06 (0.06)	-0.12 (0.02)
NXTL	-0.02 (0.01)	-0.07 (0.10)	-0.24 (0.09)	0.19 (0.03)	-0.04 (0.02)	-0.04 (0.02)
ORCL	-0.08 (0.01)	-0.21 (0.08)	-0.42 (0.03)	0.16 (0.02)	-0.03 (0.00)	-0.03 (0.01)
PSFT	0.02 (0.01)	-0.07 (0.16)	-0.08 (0.15)	0.16 (0.05)	-0.02 (0.02)	-0.07 (0.03)
QWST	-0.13 (0.01)	-0.03 (0.10)	-0.30 (0.07)	0.25 (0.03)	-0.05 (0.01)	-0.05 (0.01)
SBUX	-0.01 (0.01)	-0.07 (0.14)	-0.04 (0.06)	0.13 (0.03)	-0.11 (0.03)	-0.13 (0.02)
SUNW	-0.01 (0.05)	-0.13 (0.13)	0.43 (0.50)	0.23 (0.06)	0.00 (0.05)	0.47 (1.55)
WCOM	-0.02 (0.03)	0.08 (0.06)	0.51 (0.02)	0.25 (0.01)	0.01 (0.01)	-0.04 (0.00)

Note: This table reports the duration parameters as in (6.7) and the correlations between the innovation in the efficient price. The first three columns report the duration parameters  $\delta_1, \delta_2, \delta_3$ , with standard errors reported in brackets. The last three columns report the parameters concerning the correlations between the transitory components and the innovation in the efficient price, where  $\alpha$  concerns the transactions and  $\beta_b$  and  $\beta_a$  concerns the bid and ask quote respectively.

Table 6.3 reports the parameter estimates for the various duration functions and the correlations with the innovation in the efficient price. The first column in table 6.3 reports the results for  $\delta_1$ , the duration parameter on the innovation in the random walk. This parameter measures the effect that time between events ( $\tau_t$ ) has on the innovation in the random walk. The hypothetical values of interest are  $\delta_1 = \frac{1}{2}$  and  $\delta_1 = 0$ . A parameter value of  $\delta_1 = \frac{1}{2}$  corresponds to a price process that evolves in calendar time. Here the variance of the innovation in the random walk grows proportional to the time between events. This time scale is easily rejected for all of the stocks considered, all parameter are significantly different from this value. For the parameter value equal to 0, the results are inconclusive. Overall we find that the effect of time between events is slightly negative, 11 out of the 20 stocks report significant negative parameters, with an average parameter value of  $-0.05$ . This indicates that when the time between events is large, the innovation in the efficient price, and thus the informativeness of events, decreases. Only one stock reports a significant positive value. These findings do not conclusively reject the hypothesis that the price process evolves in our defined event time.

The second column reports the impact of time between transactions on the correlation

between the innovation in the efficient price and the microstructure noise ( $\delta_2$ ). For almost all stocks this parameter is insignificant (only one stock reports a significant parameter value). Hence time between transactions does not have a significant impact on the correlation between the innovation in the efficient price and the microstructure noise, although the average value of this parameter is -0.10.

The third column reports the parameter that measures the impact of time between quote innovations on the correlation between the innovation in the efficient price and deviations from this in quotes. As mentioned these deviations can exist due to asymmetries in information and temporary inventory imbalances. The average value for this parameter is -0.15, where 10 stocks report a significantly negative parameter value (for 3 stock we find a significantly positive value). This indicates that when the duration between quote innovations are short, there is more information in the deviations from the efficient price than at long durations.

In columns four to six we consider the correlations between the innovation in the efficient price and the corresponding microstructure noise/information asymmetry. We find that these parameters are generally positive for transactions and negative for quotes. Moreover, most of the parameters for the quotes are insignificant.

In table 6.4 we report the estimates of the parameters concerning the innovation in the random walk and the noise. Standard errors are not reported as all parameters are highly significant.

The first column reports the innovation in the random walk due to trades ( $\sigma_P$ ). We report this parameter multiplied by a factor 10. This parameter can be seen as the amount of private information per transaction. The number is inversely related to the number of transactions used in the model. The results for the innovation in the random walk due to quote innovations ( $\sigma_Q$ ) is reported in the next column. We find that the contribution to the innovation in the random walk due to public information is much larger than the innovation due to private information. Additionally, the innovation due to quotes, like the innovation due to trades, is inversely related to the liquidity of the stock. The final parameter corresponding to innovations in the random walk is reported in column 3 of table 6.4. This parameter considers the innovation in the random walk when both events occur simultaneously ( $\sigma_{PQ}$ ). When both events occur both public and private information is released. However, if only the public information at this time is considered,  $\sigma_{PQ}$  would be equal to  $\sigma_Q$ . We observe that in most cases both events are taken into account so that  $\sigma_{PQ} > \sigma_Q$ .

As we decompose trades and quotes into an informative part and noise, the model provides us with estimates on the amount of noise in trades and quotes. Column 4 and 5 in table 6.4 reports the noise-to-signal ratio for trades and quotes respectively. This ratio is defined as the variance of the "noise" over the variance of the innovation in the random walk. At



Table 6.4: Innovations in the efficient price and noise-to-signal ratios

Symbol	$\sigma_P$	$\sigma_Q$	$\sigma_{PQ}$	$NS_P$	$NS_Q$
AAPL	0.132	0.862	0.938	56.99	10.52
AMAT	0.100	0.589	0.690	55.09	13.72
AMGN	0.071	0.588	0.669	148.39	18.39
AMTD	0.203	0.993	1.124	98.82	13.26
ATHM	0.086	0.654	0.731	177.79	18.32
CIEN	0.190	1.361	1.421	73.98	6.94
COMS	0.114	1.176	1.122	121.04	21.55
CPWR	0.247	1.154	1.296	41.21	9.30
CSCO	0.048	0.372	0.417	105.26	21.18
DELL	0.059	0.711	0.707	200.95	11.93
INTC	0.057	0.444	0.445	99.91	27.95
MSFT	0.039	0.351	0.374	130.78	117.12
NOVL	0.230	1.403	1.545	33.61	6.04
NXTL	0.096	0.946	1.086	139.92	9.85
ORCL	0.109	0.919	0.984	102.01	9.42
PSFT	0.304	2.079	2.152	45.49	4.19
QWST	0.086	0.699	0.857	159.64	16.72
SBUX	0.136	1.116	1.357	105.01	9.77
SUNW	0.076	0.518	0.531	80.75	67.83
WCOM	0.035	0.436	0.471	220.69	18.08

This table reports the parameter estimates for the innovation terms in the random walk of the model in (6.7). The relevant equation that we consider here is

$$\sigma_\ell = \tau_\ell^{\delta_\ell} (P_\ell(1 - Q_\ell)\sigma_P + Q_\ell(1 - P_\ell)\sigma_Q + P_\ell Q_\ell \sigma_{PQ}).$$

*Note:* The first column reports the ticker symbol of the stock. The second column refers to the innovation in the efficient price due to transactions ( $\sigma_P$ ). The third column reports the innovations in the efficient price due to quotes ( $\sigma_Q$ ) and the fourth column reports the innovation in the efficient price due to both events occurring simultaneously ( $\sigma_{PQ}$ ). The last two columns report the noise-to-signal ratios for both trades and quotes. These are defined at average durations as

$$NS_P = \frac{\alpha^2 \sigma_P^2 + \text{Var}(u_\ell)}{\sigma_P^2}, NS_Q = \frac{\bar{\beta}^2 \sigma_Q^2 + \text{Var}(e_\ell)}{\sigma_Q^2}$$

for trades and quotes respectively. For the quotes we take the average values of the covariance with the innovation in the efficient price ( $\bar{\beta}$ ) and the average of the idiosyncratic noise ( $\text{Var}(e_\ell)$ ).

average durations this ratio is given by

$$NS_P = \frac{\alpha^2 \sigma_P^2 + \text{Var}(u_\ell)}{\sigma_P^2}, NS_Q = \frac{\bar{\beta}^2 \sigma_Q^2 + \text{Var}(e_\ell)}{\sigma_Q^2}, \quad (6.8)$$

for trades and quotes respectively. For quotes we consider the average value for the correlation parameter and the average variance over the bid and ask quote. For trades the average noise is approximately 100 times bigger than the informative part in transaction prices. However this noise differs substantially per asset, with a minimum of 34 and a maximum of 220. This

ratio can be seen as a ratio of noise traders to informed traders. As is clear from these numbers, informed traders only represent a very small fraction of the total pool of traders.

The same noise-to-signal ratio is defined for quotes, and is reported in column 5. This ratio is calculated as the average variance of the microstructure noise for bid and ask, over the variance of the innovation in the random walk due to quote innovations. For quotes this ratio can be understood as representing the uninformative component of the inside quote. This may be seen as deviations from the efficient price due to some inventory imbalances, liquidity effects and the effects of price discreteness. Clearly, the microstructure noise around quotes is proportionally much smaller than the microstructure noise around trades. The average value for this ratio is about 25, but is driven mainly by some large outliers. This confirms the notion that quotes, in most instances are less noisy than the trades (see e.g. Hansen and Lunde, 2004).

From the results concerning private and public information, we can conclude that the private information component is much smaller than the public information component. However, the results presented here provide us with the information content of single events occurring, i.e. a single transaction or a single quote innovation. A more relevant question is to what extent these information components can explain the movements in daily stock prices.

To determine the extent to which daily prices change due to private and public information, we aggregate all the innovations in the efficient price up to a daily level. For one particular day, the fraction of public/private information causing the total daily price change can be calculated by

$$PC_P = \frac{\sum_{i=0}^L \tau_\ell^{2\delta_1} (P_\ell(1 - Q_\ell)\sigma_P^2)}{\sum_{i=0}^L \tau_\ell^{2\delta_1} (P_\ell(1 - Q_\ell)\sigma_P^2 + Q_\ell(1 - P_\ell)\sigma_Q^2 + P_\ell Q_\ell \sigma_{PQ}^2)}, \quad (6.9)$$

$$PC_Q = \frac{\sum_{i=0}^L \tau_\ell^{2\delta_1} Q_\ell(1 - P_\ell)\sigma_Q^2}{\sum_{i=0}^L \tau_\ell^{2\delta_1} (P_\ell(1 - Q_\ell)\sigma_P^2 + Q_\ell(1 - P_\ell)\sigma_Q^2 + P_\ell Q_\ell \sigma_{PQ}^2)}, \quad (6.10)$$

$$PC_{PQ} = \frac{\sum_{i=0}^L \tau_\ell^{2\delta_1} P_\ell Q_\ell \sigma_{PQ}^2}{\sum_{i=0}^L \tau_\ell^{2\delta_1} (P_\ell(1 - Q_\ell)\sigma_P^2 + Q_\ell(1 - P_\ell)\sigma_Q^2 + P_\ell Q_\ell \sigma_{PQ}^2)}, \quad (6.11)$$

where  $PC_P$  is the daily price change due to private information,  $PC_Q$  the daily price change due to public information and  $PC_{PQ}$  is the daily price change due to both events.

In table 6.5 we report the fractions that each source contributes to the daily variance of the efficient price. In brackets we report the standard deviations of these fractions over the 124 trading days. A first and overall finding is that the contribution to this variance due to trades is much smaller than the variance due to quotes. On average we find a contribution of only 9.75% that can be attributed to the transaction process/private information. For the quote process/public information we find an average contribution of 64.52%. The contribution to

Table 6.5: Daily aggregated contributions to the efficient price

Symbol	% due to trades	% due to quotes	% due to both
AAPL	8.61% (2.60%)	75.15% (6.58%)	16.24% (5.56%)
AMAT	8.41% (2.09%)	60.95% (5.76%)	30.65% (4.85%)
AMGN	4.25% (1.55%)	74.48% (6.09%)	21.27% (5.33%)
AMTD	6.23% (4.02%)	66.49% (9.51%)	27.29% (9.00%)
ATHM	3.09% (2.06%)	64.92% (10.90%)	31.99% (10.07%)
CIEH	7.06% (2.35%)	76.76% (6.78%)	16.18% (5.79%)
COMS	11.88% (5.32%)	69.80% (9.00%)	18.32% (7.26%)
CPWR	17.72% (6.20%)	66.41% (7.76%)	15.88% (4.24%)
CSCO	7.46% (4.29%)	52.16% (6.19%)	40.38% (4.96%)
DELL	10.55% (4.56%)	49.77% (8.00%)	39.67% (8.24%)
INTC	12.24% (5.79%)	54.01% (6.77%)	33.75% (5.59%)
MSFT	7.55% (3.30%)	51.58% (6.82%)	40.87% (5.75%)
NOVL	18.13% (6.52%)	69.25% (8.76%)	12.62% (4.25%)
NXTL	5.36% (2.00%)	75.23% (7.16%)	19.41% (6.98%)
ORCL	18.37% (8.03%)	55.33% (9.15%)	26.30% (5.79%)
PSFT	18.27% (6.50%)	70.34% (7.04%)	11.39% (3.79%)
QWST	6.90% (6.54%)	64.95% (12.36%)	28.14% (8.43%)
SBUX	5.54% (4.13%)	77.59% (9.60%)	16.87% (6.61%)
SUNW	7.48% (2.56%)	65.00% (6.67%)	27.52% (5.89%)
WCOM	3.41% (1.22%)	67.85% (6.05%)	28.74% (5.65%)

Note: This table reports the daily aggregated contributions to the efficient price. This aggregation is done over each day of the 124 trading days. The table reports the average fractions due to trades/quotes. Standard deviations over the 124 days are given in brackets and printed in italics.

the variance due to both constitutes of the remaining part, 25.73%. Furthermore, there seems to be no real pattern in these fractions. For trades, we find a lowest value (3.41%) for the stock of MCI WorldCom and a highest value for Oracle Corporation. We find no relationship between the liquidity of the stock and the fraction that trades contribute.

## 6.5 Model Extension

In the previous section we have decomposed the information content of both trades and quotes (public and private information). However, we have not addressed the issue as to when these events are more informative. To analyze this question we consider four additional factors.

One of the most common factors tested for is whether large volume trades convey more information than low volume trades (see e.g. Easley, Kiefer and O'Hara, 1997). This is based on the notion that informed traders prefer to trade large volumes, to exploit the information that they have. On the other hand, an informed trader might wish to hide her information and split her order into a few smaller order. Barclay, Dunbar, and Warner (1993) suggest that informed traders trade at medium size volumes, as high volume trades would reveal

their information and low volume trades are to costly in terms of transaction fees, a result confirmed by Chakravarty (2001).

Additionally, we address the public information component in the market. First, we consider the total volume quoted at the inside. Large volumes quoted at the inside indicates a dealer's willingness to trade large quantities. As dealers are risk averse towards informed traders, large quoted quantities means that they expect the presence of few informed traders.<sup>7</sup>

Closely related to this is the total number of dealers at the inside. Although this measure is strongly related to quoted volume it differs in one respect. A certain amount of volume may be quoted by one single dealer or shared by many. The one dealer quoting large volume can easily withdraw her quote. When there are many dealers quoting the inside, the inside quote will be more firm. The amount of market makers at the inside is thus informative about the firmness of the inside quote.

Finally, we want to test for the type of dealer that the inside quote is coming from. Recent research (Huang, 2002) has indicated that ECNs often quote more efficient than traditional market makers. Hence, we test for the effect of whether quotes of an ECN or market maker are more informative.

To test for the impact of these factors on the different information component we construct different categories for the additional factors (high/med/low volume, many/some/few market makers at the inside and ECN/Market Makers/Both at the inside) and include these in our state space framework in (6.7) by modifying equations (6.4), (6.5) and (6.6). For the traded volume we replace  $\sigma_P$  by

$$(D_{1,\ell} \ D_{2,\ell} \ D_{3,\ell}) \begin{pmatrix} \sigma_{P,1} \\ \sigma_{P,2} \\ \sigma_{P,3} \end{pmatrix},$$

where  $D_{1,\ell}, D_{2,\ell}, D_{3,\ell}$  are dummy indicators equal to 1 if the observation falls within this category and zero otherwise. For the factor that affect quotes,  $\sigma_Q$  is replaced by

$$(D_{1,\ell} \ D_{2,\ell} \ D_{3,\ell}) \begin{pmatrix} \sigma_{Q,1} \\ \sigma_{Q,2} \\ \sigma_{Q,3} \end{pmatrix},$$

and in all cases  $\sigma_{PQ}$  is replaced by

$$(D_{1,\ell} \ D_{2,\ell} \ D_{3,\ell}) \begin{pmatrix} \sigma_{PQ,1} \\ \sigma_{PQ,2} \\ \sigma_{PQ,3} \end{pmatrix}.$$

<sup>7</sup> Additionally, Engle and Patton (2004) tests for the impact of aggregated trade volume on the innovation in dealer quotes.

Consequently a Lagrange Multiplier test can be performed to test whether the additional factors significantly improve the model. The model can then be extended in that direction.

We first test for the impact of different trading volumes. These volumes are obtained from the trade data provided by Nasdaq. To determine the impact of this volume on the innovation in the efficient price (in this case  $\sigma_P$  and  $\sigma_{PQ}$ ) we create three different categories, low volume (less than 200 shares traded), medium volume (between 200 and 1,000 shares traded) and high volume (more than 1,000 shares traded). Some summary statistics on traded volume are given in Panel A of table 6.6. The mean and standard deviation for traded volume are given in the first two columns of the table. The high mean relative to the high fraction of volume over in the medium and low categories indicates the large skewness of the distribution of traded volume. The categories, reported in the next three columns were chosen as such to have enough observations in each category.

Another factor we want to test for is whether quoted volume at the inside quote has some explanatory power. This quoted volume at the inside is obtained from the dealer quote data provided by Nasdaq, as quoted volumes are not given with the inside quotes. From the dealer quote data we continuously maintain an array of all most recent dealer quotes. These dealer quotes are then matched with the inside quotes. Every time the inside quote is innovated, we record the total volume quoted at the inside by all dealers. Likewise we record the number of dealers at the inside and the identity of the dealer (Market Maker or ECN), which are used later on.

Again three categories are made for quoted volume, where low volume quotes are below 500, medium between 500 and 2,000 and high volume above 2,000. Summary statistics on quoted volume at the inside are reported in Panel B of table 6.6. Mean and standard deviation for quoted volume are reported in the first two columns. Interesting is the fact that on average quoted volume at the inside is approximately twice the average traded volume. Additionally we observe the same pattern as in trades. In most cases the mean lies above the median, again indicating the right skewed distribution.

Related to the total volume quoted at the inside is the number of market makers at the inside. Panel A in table 6.7 reports some summary statistics on the number of dealers at the inside quote. We find that on average the number of dealers at the inside quote lies between 2 and 4 for the stocks in our sample. Again categories are made based on 1 dealer at the inside, few dealers at the inside (2-5) and many dealers at the inside ( $> 5$ ). Overall we find that half of the time a single dealer is at the inside, however there is still a considerable fraction of times that more than five dealers are at the inside.

The last factor that we consider is the type of dealer at the inside. We split up dealers in two separate categories, market makers and ECNs. Within the ECN category we include ECNs, like Island ECN and Instinet, but this category also includes the quotes disseminated from the Chicago Stock exchange. The category market makers, includes all traditional

Table 6.6: Total Traded and Quoted Volume

	Panel A					Panel B				
	Mean	Std	< 2	2 – 10	> 10	Mean	Std	< 5	5 – 20	> 20
AAPL	10.84	41.75	0.3512	0.5176	0.1311	21.55	34.31	0.2255	0.4573	0.3172
AMAT	8.63	29.13	0.3708	0.5328	0.0964	20.34	27.41	0.2451	0.4368	0.3182
AMGN	9.51	29.27	0.3955	0.4814	0.1231	14.96	17.09	0.2747	0.5065	0.2188
AMTD	3.80	7.13	0.5763	0.3835	0.0403	6.99	10.43	0.5840	0.3699	0.0461
ATHM	4.45	9.50	0.5226	0.4276	0.0498	8.95	12.35	0.4856	0.4353	0.0791
CIEI	9.29	22.89	0.2877	0.5807	0.1316	28.15	40.93	0.1702	0.4552	0.3745
COMS	9.76	44.65	0.3281	0.5429	0.1290	49.90	88.73	0.1581	0.3305	0.5115
CPWR	14.79	55.61	0.2889	0.5456	0.1655	21.03	28.83	0.2075	0.4916	0.3009
CSCO	7.38	23.57	0.4161	0.4814	0.1025	24.69	39.16	0.2248	0.4216	0.3536
DELL	8.94	34.02	0.3770	0.4889	0.1341	59.87	123.84	0.1378	0.3404	0.5218
INTC	9.62	37.52	0.3837	0.4876	0.1288	30.53	51.44	0.1914	0.4060	0.4026
MSFT	9.49	43.74	0.3910	0.4762	0.1329	34.35	58.83	0.1698	0.3979	0.4323
NOVL	13.08	48.53	0.2824	0.5322	0.1854	42.83	56.32	0.1457	0.3441	0.5102
NXTL	10.89	49.23	0.3368	0.5314	0.1318	21.79	28.81	0.2255	0.4501	0.3244
ORCL	12.61	57.33	0.3290	0.5117	0.1593	48.23	90.89	0.1480	0.3620	0.4901
PSFT	11.64	36.41	0.2880	0.5383	0.1737	45.23	70.04	0.1455	0.3648	0.4897
QWST	8.20	29.01	0.4117	0.4822	0.1061	16.83	29.35	0.2759	0.5128	0.2114
SBUX	8.15	31.58	0.4228	0.4708	0.1063	17.65	30.95	0.2945	0.4740	0.2315
SUNW	7.61	23.52	0.3882	0.5197	0.0922	19.14	26.26	0.2504	0.4641	0.2856
WCOM	9.90	35.23	0.3778	0.4886	0.1336	25.91	34.93	0.1813	0.4292	0.3895

*Note:* This table reports summary statistics on the total traded volume at the each transaction and the total volume quoted at the inside. All values should be multiplied by 100. Panel A reports statistics on the traded volume. The first column in this Panel reports the average traded volume. The second column reports the associated standard deviation. The third column reports the fraction of trades, where the traded volume is less than 200. The fourth column reports the fraction where traded volume is between 200 and 1,000. The last column reports the fraction when traded volume is more than 1,000. Panel B reports statistics on quoted volume. The first column reports the average quoted volume at the inside. The second column reports the associated standard deviation. The third column reports the fraction of quotes, where the quoted volume at the inside is less than 500. The fourth column reports the fraction where quoted volume is between 500 and 2,000 dealers are at the inside. The last column reports the fraction when quoted volume is more than 2,000.

market makers, but also those entities listed as order entry firms. The latter however only represent a very small fraction of all quotes being disseminated by market makers.

Panel B in table 6.7 reports the fractions that either ECNs solely make the inside, market makers make the inside or both make the inside. Overall the highest fraction is found for ECNs being at the inside, followed by both at the inside. The lowest fraction is found for only market makers being at the inside. Again we find that all LM-statistics are highly significant. Therefore it clearly matters, which kind of dealer is at the inside.

Consequently, we perform a Lagrange Multiplier test on the addition of the several factors in the model. In addition to the parameters estimated in (6.7) we add four additional parameters in the model (two for the innovation in the efficient price due to trades and two for the innovation due to trades and quotes). As such the LM-statistic is  $\chi^2$  distributed with 4

Table 6.7: Number and Type of Dealers at the Inside Quote

	Panel A					Panel B		
	Mean	Std	1 Dealer	2 - 5 Dealers	> 5 Dealers	ECN	Market Maker	Both
AAPL	2.58	2.63	0.5424	0.3432	0.1143	0.4115	0.2460	0.3425
AMAT	2.45	2.33	0.5329	0.3681	0.0990	0.4710	0.1642	0.3649
AMGN	2.10	2.06	0.6238	0.3058	0.0704	0.4699	0.2759	0.2542
AMTD	1.36	0.92	0.7909	0.1997	0.0095	0.6245	0.2147	0.1608
ATHM	1.49	1.17	0.7454	0.2377	0.0169	0.5796	0.2289	0.1915
CIEN	2.83	3.03	0.5254	0.3325	0.1421	0.3773	0.2581	0.3645
COMS	4.34	4.81	0.4596	0.2500	0.2904	0.3501	0.1894	0.4605
CPWR	2.46	2.51	0.5697	0.3219	0.1085	0.3915	0.3082	0.3003
CSCO	2.53	2.57	0.5253	0.3697	0.1050	0.4774	0.1465	0.3762
DELL	4.14	4.59	0.4400	0.2905	0.2695	0.3616	0.1573	0.4811
INTC	2.90	3.00	0.4980	0.3508	0.1512	0.4211	0.1741	0.4049
MSFT	2.70	2.76	0.5125	0.3624	0.1251	0.4609	0.1468	0.3923
NOVL	4.26	4.51	0.4678	0.2303	0.3019	0.3353	0.2507	0.4140
NXTL	2.60	2.71	0.5473	0.3357	0.1169	0.4057	0.2656	0.3287
ORCL	4.09	4.58	0.4662	0.2669	0.2668	0.3500	0.2036	0.4464
PSFT	4.12	4.40	0.4830	0.2315	0.2856	0.3169	0.2741	0.4090
QWST	2.09	2.25	0.6252	0.3081	0.0667	0.4565	0.2883	0.2552
SBUX	2.46	2.72	0.5917	0.2976	0.1107	0.4061	0.3214	0.2725
SUNW	2.34	2.29	0.5568	0.3542	0.0890	0.4746	0.1816	0.3438
WCOM	2.62	2.70	0.5189	0.3705	0.1106	0.4320	0.1931	0.3749

*Note:* This table reports summary statistics on the number of dealers at the inside quote (both market makers and ECNs) and the type of dealers at the inside (Market Maker or ECN). Panel A reports some statistics on the number of dealers at the inside. The first column in this Panel reports the average number of dealers at the inside. The second column reports the associated standard deviations. The third column reports the fraction of quotes, where a single dealer is at the inside. The fourth column reports the fraction, when between 2 and 5 dealers are at the inside. The last column reports the fraction where more than 5 dealers are at the inside. Panel B reports statistics on the type of dealer at the inside. The first column in this Panel reports the fraction of time ECNs are at the inside alone (this includes the quotes disseminated by the Chicago Stock Exchange), the second column reports the fraction of time market makers are the only ones setting the inside and the last column is the fraction when both set the inside.

degrees of freedom. The critical values are 9.49 and 13.3 at the 5% and 1% level respectively.

The LM-statistics for traded volume are reported in the first column of table 6.8. For half of the stocks we find that the inclusion of the additional parameters will have a significant impact on the likelihood of (6.7). These results are in line with Easley, Kiefer, and O'Hara (1997) who find that traded volume in some cases leads to a significant improvement in the likelihood function.

LM-statistics for quoted volume are report in the second column of table 6.8. All statistics are highly significant and indicate improvements of the likelihood function in this direction. Therefore quoted volume is much more informative about the amount of public information included in quotes.

Column 3 in table 6.8 reports the LM-statistics for the number of dealers at the inside.

Table 6.8: Lagrange Multiplier Statistics

	Traded Vol	Quoted Vol	# of MM/ECN	MM/ECN
AAPL	8.75	433.1	578.5	448.4
AMAT	3.41	729.9	1284.3	1225.7
AMGN	6.33	359.6	709.4	445.7
AMTD	66.24	97.3	231.5	438.4
ATHM	13.60	216.2	488.6	437.2
CIEN	6.66	346.1	534.7	405.6
COMS	8.29	60.8	90.7	201.2
CPWR	33.93	622.8	862.0	611.5
CSCO	28.69	1947.6	2813.6	1388.0
DELL	37.07	784.7	1340.9	1154.0
INTC	12.46	647.9	998.1	862.8
MSFT	65.67	137.4	226.4	296.0
NOVL	25.25	138.9	172.2	155.1
NXTL	9.38	209.7	410.7	218.1
ORCL	28.42	1115.3	1511.9	1204.0
PSFT	12.69	120.6	206.4	202.4
QWST	41.59	906.2	1008.0	433.4
SBUX	0.44	222.2	313.3	200.3
SUNW	6.12	274.5	484.2	496.0
WCOM	17.92	95.5	508.8	397.3

*Note:* This table reports the Lagrange Multiplier statistics for the additional factors tested for. These factors include transaction volume, total quoted volume at the inside quote. The total number of market makers/ECNs at the inside quote and the fact whether market makers or ECN set the inside. Results are reported in columns 1 to 4 respectively. In all cases the LM-statistic is  $\chi^2$  distributed with 4 degrees of freedom. The critical values are 9.49 and 13.3 at the 5% and 1% level respectively.

Again all statistics are highly significant. Compared to the total volume quoted at the inside we find that the LM-statistic for the number of dealers at the inside is higher in all cases, hence more dealers at the inside is more informative than the quoted volume at the inside.

The fourth column of table 6.8 reports the LM-statistics for the type of dealer at the inside. All statistics are highly significant, indicating that the type of dealer at the inside matters.

The LM-statistics discussed in above give us a clear indication to which factors most affect the innovation in the efficient price. For the trades we found significant LM-statistics in only 50% of the cases. Hence, we do not expand the model in this direction. For quoted volume at the inside we found all LM-statistics to be significant, however we expect greater improvement for the number of dealers at the inside. This is one direction in which we expand the model. Finally, we expand the model, incorporating information as to who is at the inside, market makers or ECNs. This direction is of interest given recent literature (Huang, 2002) which finds that ECNs in general lead to a contribution to the price discovery process.

We estimate (6.7) with the inclusion of the dummy function for both the number of dealers



Table 6.9: Number of Dealers at the inside

	1 Dealer	2 - 5 Dealers	> 5 Dealers	1 Dealer	2 - 5 Dealers	> 5 Dealers
	$\sigma_Q, Few$	$\sigma_Q, Some$	$\sigma_Q, Many$	$\sigma_{PQ}, Few$	$\sigma_{PQ}, Some$	$\sigma_{PQ}, Many$
AAPL	0.75 (0.01)	0.95 (0.01)	1.09 (0.02)	0.87 (0.02)	0.99 (0.05)	1.09 (0.03)
AMAT	0.50 (0.00)	0.65 (0.01)	0.77 (0.01)	0.64 (0.03)	0.73 (0.01)	0.77 (0.02)
AMGN	0.52 (0.01)	0.65 (0.01)	0.82 (0.01)	0.62 (0.02)	0.71 (0.04)	0.81 (0.03)
AMTD	0.96 (0.01)	1.10 (0.02)	1.32 (0.03)	1.11 (0.02)	1.14 (0.04)	1.33 (0.05)
ATHM	0.62 (0.01)	0.72 (0.01)	0.93 (0.02)	0.71 (0.01)	0.76 (0.02)	0.87 (0.03)
CIEN	1.17 (0.02)	1.49 (0.02)	1.71 (0.02)	1.36 (0.04)	1.44 (0.04)	1.63 (0.05)
COMS	1.08 (0.06)	1.19 (0.03)	1.30 (0.02)	1.13 (0.08)	1.12 (0.07)	1.12 (0.04)
CPWR	0.94 (0.02)	1.28 (0.02)	1.66 (0.04)	1.17 (0.05)	1.38 (0.09)	1.61 (0.07)
CSCO	0.31 (0.00)	0.39 (0.00)	0.51 (0.00)	0.36 (0.01)	0.43 (0.01)	0.48 (0.01)
DELL	0.50 (0.01)	0.65 (0.01)	0.96 (0.01)	0.58 (0.02)	0.66 (0.04)	0.95 (0.02)
INTC	0.35 (0.07)	0.46 (0.02)	0.61 (0.02)	0.40 (0.08)	0.45 (0.02)	0.61 (0.03)
MSFT	0.36 (0.01)	0.41 (0.11)	0.42 (0.01)	0.25 (0.02)	0.21 (0.04)	0.29 (0.02)
NOVL	1.18 (0.05)	1.54 (0.03)	1.63 (0.04)	1.41 (0.06)	1.77 (0.12)	1.56 (0.07)
NXTL	0.83 (0.01)	1.02 (0.02)	1.20 (0.02)	1.06 (0.05)	1.06 (0.06)	1.25 (0.06)
ORCL	0.68 (0.03)	0.85 (0.01)	1.26 (0.01)	0.80 (0.02)	0.96 (0.04)	1.27 (0.02)
PSFT	1.75 (0.07)	2.36 (0.05)	2.46 (0.04)	2.02 (0.10)	2.29 (0.11)	2.37 (0.10)
QWST	0.61 (0.01)	0.76 (0.01)	1.06 (0.02)	0.80 (0.02)	0.88 (0.03)	1.10 (0.05)
SBUX	0.97 (0.02)	1.24 (0.03)	1.48 (0.03)	1.32 (0.19)	1.32 (0.06)	1.61 (0.07)
SUNW	0.46 (0.05)	0.54 (0.03)	0.68 (0.02)	0.50 (0.02)	0.54 (0.03)	0.67 (0.03)
WCOM	0.39 (0.02)	0.46 (0.00)	0.54 (0.01)	0.44 (0.01)	0.49 (0.01)	0.53 (0.02)

Note: This table reports the results for the extension in model (6.7), where we allow the innovation in the efficient price to depend on the number of dealers that are at the inside, or Electronic Communication Networks. The first three columns refer to innovations in the quotes, with low, medium or a high number of market makers at the inside. The last three to innovations in both quotes and trades.

at the inside and the type of dealer.

Table 6.9 reports the results for the number of market makers at the inside. As the number of market makers at the inside only affects the parameters  $\sigma_Q$  and  $\sigma_{PQ}$  only these parameters are reported. Further, the parameter  $\sigma_P$  is not much affected by the inclusion of the dummy variables. There is a very clear trend for all the stocks in our sample. The more dealers at the inside quote the larger the effect a quote innovation has on the efficient price. Hence the more dealers are at the inside the more informative these quotes are (as the fraction of  $\sigma_Q/\sigma_P$  increases). Another interesting feature observed from table 6.9 is are the differences observed between  $\sigma_Q$  and  $\sigma_{PQ}$ . When there are few dealers at the inside, trades clearly add information to the innovation in the efficient price (for the low category  $\sigma_{PQ}$  is in almost all cases larger than  $\sigma_Q$ ). However, when there are many dealers at the inside, the additional information that trades add disappears (in some cases we even observe lower values for  $\sigma_{PQ}$ ).

Table 6.10 reports the results for the innovations in the efficient price when different types of market makers are at the inside. Again only the results are reported for  $\sigma_Q$  and  $\sigma_P$ . The effect of the inclusion of these dummy variables again has only a small impact on

Table 6.10: Market Makers or ECN at the inside

	$\sigma_{Q,ECN}$	$\sigma_{Q,MM}$	$\sigma_{Q,Both}$	$\sigma_{PQ,ECN}$	$\sigma_{PQ,MM}$	$\sigma_{PQ,Both}$
AAPL	0.73 (0.01)	0.93 (0.01)	0.97 (0.01)	0.82 (0.03)	1.05 (0.06)	0.97(0.08)
AMAT	0.49 (0.01)	0.68 (0.03)	0.69 (0.01)	0.59 (0.02)	0.71 (0.03)	0.73(0.01)
AMGN	0.51 (0.01)	0.64 (0.01)	0.68 (0.01)	0.60 (0.02)	0.71 (0.03)	0.70(0.06)
AMTD	0.89 (0.01)	1.16 (0.02)	1.16 (0.02)	1.00 (0.04)	1.15 (0.05)	1.05(0.08)
ATHM	0.61 (0.01)	0.70 (0.01)	0.78 (0.01)	0.66 (0.02)	0.70 (0.02)	0.73(0.03)
CIEN	1.12 (0.02)	1.44 (0.02)	1.55 (0.02)	1.29 (0.06)	1.43 (0.08)	1.39(0.05)
COMS	1.05 (0.02)	1.12 (0.12)	1.22 (0.02)	1.48 (0.10)	1.51 (0.09)	1.41(0.11)
CPWR	0.88 (0.02)	1.21 (0.02)	1.41 (0.03)	1.16 (0.07)	1.29 (0.15)	1.31(0.07)
CSCO	0.32 (0.00)	0.38 (0.01)	0.44 (0.00)	0.36 (0.01)	0.42 (0.01)	0.45(0.01)
DELL	0.49 (0.01)	0.67 (0.03)	0.86 (0.01)	0.53 (0.02)	0.70 (0.08)	0.82(0.04)
INTC	0.43 (0.01)	0.44 (0.12)	0.58 (0.01)	0.29 (0.02)	0.29 (0.13)	0.35(0.04)
MSFT	0.36 (0.01)	0.41 (0.11)	0.42 (0.01)	0.25 (0.02)	0.21 (0.05)	0.29(0.02)
NOVL	1.12 (0.03)	1.51 (0.07)	1.56 (0.03)	1.42 (0.15)	1.71 (0.16)	1.58(0.14)
NXTL	0.81 (0.02)	1.02 (0.02)	1.06 (0.02)	0.93 (0.07)	1.12 (0.06)	1.06(0.08)
ORCL	0.67 (0.01)	0.87 (0.04)	1.13 (0.01)	0.79 (0.03)	0.88 (0.04)	1.07(0.03)
PSFT	1.67 (0.05)	2.10 (0.08)	2.40 (0.04)	1.90 (0.14)	2.17 (0.16)	2.18(0.13)
QWST	0.58 (0.01)	0.76 (0.01)	0.87 (0.01)	0.74 (0.03)	0.90 (0.06)	0.90(0.04)
SBUX	0.94 (0.02)	1.19 (0.05)	1.30 (0.04)	1.17 (0.08)	1.27 (0.08)	1.30(0.10)
SUNW	0.49 (0.03)	0.62 (0.20)	0.64 (0.08)	0.31 (0.03)	0.35 (0.08)	0.38(0.04)
WCOM	0.38 (0.01)	0.48 (0.04)	0.49 (0.01)	0.42 (0.02)	0.51 (0.02)	0.48(0.02)

Note: This table reports the results for the extension in model (6.7), where we allow the innovation in the efficient price to depend on whether traditional market makers are at the inside, or Electronic Communication Networks. The first three columns refer to innovations in the quotes, the last three to innovations in both quotes and trades.

the innovation in the efficient price due to trades. In all cases we find that market makers are more informative at the inside than ECNs. This can be due to the fact that ECNs are typically the more efficient dealers (see Huang, 2002) and at the inside most of the time. Then, when a market maker sets the inside quote this might be more informative about the location of the efficient price.

Another result is that quotes are most informative when both ECNs and market makers are at the inside, confirming the result that more dealers at the inside reveals more information. We finally note that when either an ECN or a market maker is at the inside, trades still add some additional information ( $\sigma_{PQ}$ ). However, when both are at the inside the additional information of trades disappears. This is of course related to the same result obtained for the number of market makers at the inside. When both are at the inside, there are at least 2 dealer that have this quote.

### 6.6 Conclusion

In this chapter we proposed a model for the decomposition of public and private information. This model captures the extend to which public information affects the efficient price and

to what extent private information moves the efficient price. The public information is obtained from the quote data. Private information is obtained from the trade data. We find, on average, that 10% of daily stock price movements can be explained by private information, which is in line with the findings of Madhavan, Richardson, and Roomans (1997). As the model uses all available data when it arrives to the market, we can additionally test for the impact of time on the innovation in the efficient price. Here we find that time has little impact on the innovation in the efficient price, confirming the findings of Clark (1973).

We extend the model to include several other relevant microstructure variables. We test for the impact that trade volume has on the private information component and find that this leads to only a small improvement in the likelihood function. On the contrary, the impact of quoted volume at the inside has on the public information component leads to a very significant improvement in the likelihood of the model. Even more informative are the amount of dealers that simultaneously are at the inside. When many dealers are at the inside, inside quotes are more informative. We also find that the type of dealer at the inside matters. Traditional market makers at the inside are more informative than ECNs at the inside. However, the inside quotes are most informative when both are at the inside, which confirms the notion that multiple market makers at the inside are more informative.

## Chapter 7

# Summary and Concluding Remarks

This thesis has focused on several empirical market microstructure issues. Market microstructure, which addresses the frictions present in financial market, is important for both investors as well as regulators. As frictions lead to costs in trading, investors benefit when a market operates with the least possible frictions. Regulators, on the other hand, should ensure that a market operates with the least frictions possible. This ensures a transparent market where trade is conducted fairly and no particular trading parties have advantages over other trading parties.

The recent growth in communication networks and information technology has caused many changes in the operations of financial markets. Many financial markets worldwide are switching from the traditional floor-based trading to electronic trading. For the Nasdaq, the financial market this thesis centers around, this has led to a change in trading rules and the existence of many alternative trading venues. Ensuring that information is revealed swiftly to all venues and that the flow of information is transparent is essential for such a market. On the other hand, when information is not disseminated swiftly, markets become fragmented. This fragmentation is a source of frictions instigating a risk, which leads to an additional cost for investors.

Together with the growth in information technology, also the recording of information has increased. At the moment, many large financial markets around the world record all transactions and quotes that occur in the market and make this information accessible. As most of the microstructure effects manifest themselves at the transaction or quote level, using as much information as possible is vital. In this study, we have strived to incorporate as much of this information as possible.

The empirical chapters of this thesis are all put into a single framework in chapter 2 and address several aspects of the flow in information. Firstly, we address the question as to how and how quickly dealers respond to changes in the quotes of other dealers. Secondly, we consider the effect that dealers have on the value of the asset and how informative the

quotes of dealers are. Thirdly, we address the frictions present in the market. By removing these frictions from observed prices, we find the *true* flow of information, which results in a measure for the volatility of an asset. Lastly, we address the informational asymmetry between dealers and traders.

The remainder of this chapter addresses the empirical chapters of this thesis and the general conclusions that we find.

In the third chapter of this thesis we propose a nonlinear error correction model for the interaction between dealer quotes. Previous research has suggested that quotes error correct towards the mid-quote, which is the average of all quotes issued by all dealers. In the model we propose dealer quotes correct towards the inside quotes (the best bid and ask quote issued by all dealers). We find that this error correction term is significantly better than the correction towards the mid-quote and that all dealers correct strongly towards the inside quotes. An additional factor included in the model is an indicator, which indicates whether a particular dealer is at the inside. This indicator reveals whether a particular dealer is considered more or less informed than the other dealers. We find this term to be highly significant in the model, where we find the strongest reaction to Island ECN being at the inside. More specifically, this term reveals that except for the Island ECN all dealers want to move away from the inside, when they are at the inside. Also, all dealers respond to Island being at the inside, by moving towards the inside as well. Additionally, all dealers are inclined to keep spreads small, but prefer to do this by raising their bid, instead of lowering their ask. We further analyze dealer efficiency using impulse response functions. As the model is nonlinear, we use generalized impulse responses. These generalized impulse responses can be sensitive to the history of the model and the size of the shock applied (the linear case is insensitive to all these factors). We find that the history of the model does not influence the outcome of the impulse responses, but that the most important factor is whether the nonlinearities are included. The results of the impulse responses again indicate the importance of the Island ECN.

In the fourth chapter we propose an alternative method for addressing the price discovery issue. The traditional measure for price discovery is defined as the amount of variance that each venue or dealer contributes to the variance of the efficient price. This approach, however, has the fallacy that it cannot uniquely decompose this price contribution to each dealer. This fallacy is mainly driven by the fact that data needs to be aggregated for the traditional approach. The model we propose is designed for data that is sampled at the highest possible frequency. We therefore have a model in tick time, which also measures price discovery per tick. When addressing the price discovery issue we can define three measures for price discovery, one being the dealer's reactions to changes in the efficient price, a second being the change in the efficient price due to changes in a dealer's quote and a third being the amount of variance each dealer contributes to the variance of the efficient price. We show

that this last measure is a composition of the other two measures. All these measures are functions of the duration between quote innovations. Overall we find that quotes are more informative and the price discovery is higher at short durations. We also find that different dealers often dominate for the three different measures for price discovery. When aggregating these measures of price discovery, we find that all measures converge to one single measure, which resembles the traditional measure for price discovery. However, in this case the more data is aggregated, the more this measure will converge to a single point.

In chapter 5 we address the issue of realized volatility. Realized volatility is computed by summing squared intra-day returns and provides a consistent measure for the daily volatility of an asset. The strength of realized volatility lies in the fact that it is observed rather than latent, such as the volatility estimated with a GARCH model. However, when sampling at very high frequencies, the presence of microstructure noise contaminates the price process and makes realized volatility a biased measure of the *true* volatility. In chapter 5 we employ several corrections to high frequency returns and address the two different time scales over which data can be aggregated, calendar time and tick time. When fitting a structural model to tick time data we find that it removes all of the negative autocorrelation induced by microstructure noise and find a lower bound on the realized volatility. A recently introduced bias correction yields similar results. Sampling in tick time further reveals that there is a long persistence in intra-day tick returns. This momentum in returns is not observed when sampling is done in calendar time. Calendar time sampling only reveals negative autocorrelations and hides the momentum observed in tick returns. Correcting for the persistence in tick returns leads to a realized volatility that is consistent with realized volatility derived at lower frequencies. However, it remains questionable whether this persistence reflects a source of noise or is inherent to the price process.

Lastly, the sixth chapter addresses the information content of both trades and quotes. With the widely accepted notion that market makers have access to the public information set, the additional information that flows from trades, reveals the private information present in the market. These two information components are extracted from both trades and quotes by comparing both processes. As both trades and quotes concern the same asset, they follow the same “efficient” price process. However, trades can only occur after quotes have been set. To determine the innovation both trades and quotes bring to the price process, we allow the innovation in the efficient price to depend on whether a trade or a quote has been observed. Overall, we find that trades (private information) cause about 10% of daily price movements. Additional tests on the private and public information components reveal that there is only a weak relationship between private information and traded volume. A strong relationship is found between the quoted volume and public information, but an even stronger relationship between the number of dealers that jointly set the best quote in the market and the public information. We finally find that more public information is present when a traditional

market maker sets the best price in the market, than when an Electronic Communication Network sets the best quote.

# Bibliography

- Aït-Sahalia, Yacine, Per Mykland, and Lan Zhang, 2003, How often to sample a continuous-time process in the presence of market microstructure noise, Working Paper.
- Amihud, Yakov, and Haim Mendelson, 1980, Dealership markets: Market making with inventory, *Journal of Financial Economics* 8, 31–53.
- Andersen, Torben G., 2004, Discussion of power and bipower variation, *Journal of Financial Econometrics* 2, 37–48.
- , and Tim Bollerslev, 1998, Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review* 39, 885–905.
- , Francis X. Diebold, and Paul Labys, 2001, The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* 96, 42–55.
- Andreou, Elena, and Eric Ghysels, 2002, Rolling-sample volatility estimators: Some new theoretical, simulation and empirical results, *Journal of Business and Economic Statistics* 20, 363–376.
- Ané, Thierry, and Hélyette Geman, 2000, Order flow, transaction clock, and normality of asset returns, *Journal of Finance* 55, 2259–2284.
- Bagehot, W., 1971, The only game in town, *Financial Analyst Journal* 27, 12–14.
- Baillie, Richard T., Geoffrey Booth, Yiuman Tse, and Tatyana Zabolina, 2002, Price discovery and common factor models, *Journal of Financial Markets* 5, 309–321.
- Bandi, Frederico M., and Jeffrey R. Russell, 2003, Microstructure noise, realized volatility and optimal sampling, Mimeo Working Paper, University of Chicago.
- , 2004, Separating microstructure noise from volatility, Mimeo Working Paper.
- Barclay, Michael J., William G. Christie, Jeffrey H. Harris, Eugene Kandel, and Paul H. Schultz, 1999, Effects of market reform on the trading costs and depths of nasdaq stocks, *Journal of Finance* 54, 1–34.



- Barclay, Michael J., C.G. Dunbar, and J.B. Warner, 1993, Stealth and volatility: which trades move prices?, *Journal of Financial Economics* 34, 281–306.
- Barndorff-Nielsen, Ole E., and Neil Shephard, 2002, Estimating quadratic variation using realized volatility, *Journal of Applied Econometrics* 17, 457–477.
- Beveridge, Stephen, and Charles R. Nelson, 1981, A new approach to decomposition of economic time series into permanent and transitory components with a particular attention to measurement of the business cycle, *Journal of Monetary Economics* 7, 151–174.
- Blume, Marshal E., and Michael A. Goldstein, 1997, Quotes, order flow, and price discovery, *Journal of Finance* 52, 221–244.
- Breusch, T.S., and A. R. Pagan, 1979, A simple test for heteroskedasticity and random coefficient variation, *Econometrica* 47, 1287 – 1294.
- Chakravarty, Sugato, 2001, Stealth-trading: Which traders' trades move stock prices?, *Journal of Financial Economics* 61, 289–307.
- Chan, K.C., William G. Christie, and Paul H. Schultz, 1995, Market structure and the intraday pattern of bid-ask spreads for nasdaq securities, *Journal of Business* 68, 35–60.
- Christie, William G., Jeffrey H. Harris, and Paul H. Schultz, 1994, Why did nasdaq market makers stop avoiding odd-eighth quotes?, *Journal of Finance* 49, 1841–1860.
- Christie, William G., and Paul H. Schultz, 1994, Why do nasdaq market makers avoid odd-eighth quotes?, *Journal of Finance* 49, 1813–1840.
- Chung, Kee H., and Seong-Yeon Cho, 2004, Security analysis and market making, *Journal of Financial Intermediation* 0000, 35–60.
- Clark, P.K., 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41, 135–156.
- Coughenour, Jay, and Kuldeep Shastri, 1999, Symposium on market microstructure: A review of empirical research, *The Financial Review* 34, 1 – 28.
- De Jong, Frank, Ronald Mahieu, and Peter Schotman, 1998, Price discovery in the foreign exchange market: an empirical analysis of the yen/dmark rate, *Journal of International Money and Finance* 17, 5–27.
- De Jong, Frank, and Peter Schotman, 2003, Price discovery in fragmented markets, Working Paper.

- De La Vega, Joseph, 1688, *Confusion de confusions: Portions descriptive of the amsterdam stock exchange*, Translation by H. Kellenbenz, Harvard University (1957).
- Demsetz, Harold, 1968, The cost of transacting, *Quarterly Journal of Economics* 82, 33–53.
- Dufour, Alfonso, and Robert E. Engle, 2000, Time and the price impact of a trade, *Journal of Finance* 55, 2467–2498.
- Durbin, J., and S. J. Koopman, 2001, *Time Series Analysis by State Space Methods* (Oxford University Press: Oxford).
- Easley, David, S. Hvidkjaer, and M. O'Hara, 2002, Is information risk a determinant of asset returns, *Journal of Finance* 58, 2185–2210.
- Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1997, The information content of the trading process, *Journal of Empirical Finance* 4, 159–186.
- , and Joseph B. Paperman, 1996, Liquidity, information, and infrequently traded stocks, *Journal of Finance* 51, 1405–1436.
- Easley, David, and Maureen O'Hara, 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics* 19, 69–90.
- , 1992, Time and the process of security price adjustment, *Journal of Finance* 47, 905–927.
- Ellis, Katrina, Roni Michaely, and Maureen O'Hara, 2002, The making of a dealer market: From entry to equilibrium in the trading of Nasdaq stocks, *Journal of Finance* 62, 2289–2316.
- Engle, Robert, 2000, The econometrics of ultra-high frequency data, *Econometrica* 68, 1–22.
- Engle, Robert F., and Andrew J. Patton, 2004, Impacts of trades in an error-correction model of quote prices, *Journal of Financial Markets* 7, 1–25.
- Engle, Robert F., and Jeffrey R. Russell, 1997, Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model, *Journal of Empirical Finance* 4, 187–212.
- Frijns, Bart P. M., and Peter C. Schotman, 2004, Price discovery in tick time, CEPR Discussion Paper Nr. 4456.
- Garman, Mark, 1976, Market microstructure, *Journal of Financial Economics* 13, 71–100.

- Glosten, Lawrence, and Paul Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 3, 257 – 275.
- Glosten, L. R., and L.E. Harris, 1988, Estimating the components of the bid/ask spread, *Journal of Financial Economics* 21, 123 – 142.
- Godek, Paul, 1996, Why nasdaq market makers avoid odd eights quotes, *Journal of Financial Economics* 41, 465 – 474.
- Hamilton, James D., 1994, *Time Series Analysis* (Princeton University Press: New Jersey).
- Hansen, Peter R., and Asger Lunde, 2004a, Realized variance and iid market microstructure noise, Working Paper, Brown University.
- , 2004b, An unbiased measure of realized variance, Working Paper, Brown University.
- Harris, Jeffrey H., and Paul H. Schultz, 2002, The trading profits of SOES bandits, *Journal of Financial Economics* 50, 39–62.
- Harris, Lawrence, 1987, Transaction data tests of the mixture of distributions hypothesis, *Journal of Financial and Quantitative Analysis* 22, 127–141.
- Harvey, A., 1989, *Forecasting, structural time series models and the Kalman Filter* (Cambridge University Press: Cambridge).
- Hasbrouck, Joel, 1991, Measuring the Information content of stock trades, *Journal of Finance* 46, 179–207.
- , 1993, Assessing the quality of a security market: A new approach to transaction-cost measurement, *Review of Financial Studies* 6, 191–212.
- , 1995, One security, many markets: Determining the contribution to price discovery, *Journal of Finance* 50, 1175–1199.
- , 1999, The dynamics of discrete bid and ask quotes, *Journal of Finance* 54, 2109–2142.
- , and Thomas S. Y. Ho, 1987, Order arrival, quote behavior, and the return-generating process, *Journal of Finance* 42, 1035–1048.
- Huang, Roger D., 2002, The quality of ECN and Nasdaq market maker quotes, *Journal of Finance* 57, 1285–1319.

- , and Hans R. Stoll, 1996, Dealer vs. auction markets: A paired comparison of execution costs on Nasdaq and the NYSE, *Journal of Financial Economics* 41, 313 – 357.
- Jang, Hasung, and P.C. Venkatesh, 1991, Consistency between predicted and actual bid-ask quote revisions, *Journal of Finance* 46, 433–446.
- Kandel, Eugene, and Leslie M. Marx, 1997, Payments for order flow on nasdaq, *Journal of Finance* 54, 35–66.
- Koop, Gary, M. Hashem Pesaran, and Simon M. Potter, 1996, Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* 74, 119–147.
- Koopman, Siem Jan, and J. Durbin, 2000, Fast filtering and smoothing for multivariate state space models, *Journal of Time Series Analysis* 21, 281 – 296.
- Kyle, Albert S., 1985, Continuous auctions and insiders trading, *Econometrica* 53, 1315–1335.
- Lee, C., and M. Ready, 1991, Inferring the trade direction for intradaily data, *Journal of Finance* 46, 733 – 746.
- Lehmann, Bruce N., 2002, Some desiderata for the measurement of price discovery across markets, *Journal of Financial Markets* 5, 259 – 276.
- Lyons, Richard K., 2001, *The Microstructure Approach to Exchange Rates* (MIT Press: Cambridge, Massachusetts).
- Madhavan, Ananth, 2000, Market microstructure: A survey, *Journal of Financial Markets* 3, 205 – 258.
- , Matthew Richardson, and Mark Roomans, 1997, Why do security prices change? a transaction level analysis of NYSE stocks, *Review of Financial Studies* 10, 1035 – 1064.
- Meddahi, Nour, 2002, A theoretical comparison between integrated and realized volatility, *Journal of Applied Econometrics* 17, 479 – 508.
- O'Hara, Maureen, 1995, *Market Microstructure Theory* (Blackwell Publishers Inc.: Massachusetts).
- Øksendahl, Bernt, 1998, *Stochastic Differential Equations: An Introduction with Applications* (Springer-Verlag: Berlin, Heidelberg, New York).
- Oomen, Roel C. A., 2002, Modeling realized variance when returns are serially correlated, Working Paper, University of Warwick.

- Revuz, Daniel, and Marc Yor, 1999, *Continuous Martingales and Brownian Motion* (Springer Verlag: Berlin, Heidelberg, New York).
- Roll, Richard, 1984, A simple implicit measure of the effective bid ask spread in an efficient market, *Journal of Finance* 39, 1127–1139.
- Schultz, Paul H., 2003, Who makes markets, *Journal of Financial Markets* 6, 49–72.
- Smith, Jeffrey W., James P. Selway III, and D. Timothy McCormick, 1998, The nasdaq stock market: Historical background and current operation, NASD Working Paper 98-01.
- Stoll, Hans R., 1989, Inferring the components of the bid-ask spread: Theory and empirical tests, *Journal of Finance* 44, 115–134.
- , 2000, Friction, *Journal of Finance* 55, 1479–1514.
- Wasserfallen, Walter, and Heinz Zimmermann, 1985, The behavior of intraday exchange rates, *Journal of Banking and Finance* 9, 55–72.
- Weston, James P., 2000, Competition on the nasdaq and the impact of recent market reforms, *Journal of Finance* 55, 2565–2598.
- White, Halbert, 1982, Maximum likelihood estimation of misspecified models, *Econometrica* 50, 1–25.
- Zhang, Lan, Per Mykland, and Yacine Aït-Sahalia, 2003, A tale of two time scales: Determining integrated volatility with noisy high-frequency data, Working Paper, University of Chicago.

# Nederlandse Samenvatting

Dit proefschrift richt zich op verscheidene empirische aspecten van de microstructuur van financiële markten. Het onderzoeksgebied dat zich richt op de microstructuur van markten, bestudeert de fricties die bij het handelsproces optreden, en is van belang voor zowel investeerders als wel regelgevers. Aangezien fricties leiden tot handelskosten, zijn investeerders erbij gebaat dat markten met zo min mogelijk fricties opereren. Regelgevers, daarentegen, dienen ervoor te zorgen dat deze fricties minimaal zijn om zodoende een transparante markt te creëren, waarbij de handel eerlijk verloopt en er geen handelspartijen zijn die bevoordeeld zijn ten opzichte van andere partijen.

De recente groei in communicatienetwerken en informatie technologie heeft veel veranderingen in de manier waarop financiële markten opereren teweeggebracht. Voor veel beurzen over de gehele wereld heeft dit geleid tot een overgang van handel op de vloer, naar handel via elektronische systemen. Voor de Amerikaanse beurs, de Nasdaq, de financiële markt waar deze studie zich op richt, heeft dit geleid tot een verandering in de regelgeving omtrent handel en het bestaan van verschillende alternatieve handelssystemen. Het is voor een dergelijke markt essentieel dat de beschikbare informatie snel voor al de verschillende handelssystemen beschikbaar is, zodat de markt transparant is. Als de informatie zich niet snel over deze systemen verspreidt, dan leidt dit tot een gefragmenteerde markt. Deze fragmentatie leidt tot onzekerheid omtrent de correcte waarde van een aandeel en veroorzaakt dus een risico, hetgeen leidt tot additionele kosten voor de belegger.

Gepaard gaande met de groei in informatie technologie, groeit ook de hoeveelheid informatie die wordt opgeslagen. Veel van de grotere financiële markten slaan tegenwoordig alle gebeurtenissen op die op de markt plaats vinden (transacties en het zetten van bied en laat koersen) en maken deze informatie openbaar beschikbaar. Het is vitaal om zoveel mogelijk van deze informatie te gebruiken, aangezien de meeste fricties in het handelsproces optreden op een transactie- of prijszettingniveau. In deze studie streven wij erna om zoveel mogelijk van deze informatie te gebruiken.

De empirische hoofdstukken van dit proefschrift zullen nu nader worden toegelicht. Deze hoofdstukken volgen na hoofdstuk 2 dat een inleiding geeft tot de empirische hoofdstukken en deze in één enkel raamwerk plaatst.

In het derde hoofdstuk stellen wij een model voor om de interactie tussen dealers te beschrijven. Voorgaand onderzoek veronderstelt dat de koersen die dealers zetten de neiging hebben zich te corrigeren naar het gemiddelde van alle koersen. In ons model stellen wij een prijscorrectie naar de beste bied en de beste laat koers voor. Een correctie verder dan deze punten heeft geen zin. Wij vinden dat deze error-correctie term de dynamiek van koersen significant beter beschrijft dan een correctie naar het middenpunt en dat dealers een sterke correctie hebben naar deze koersen. Een additionele factor die we gebruiken is een indicator die aangeeft of een bepaalde dealer de beste bied of laat koers zet. Als de koersen van andere dealers significant reageren op het feit dat een bepaalde dealer de beste koers(en) zet, dan wordt deze dealer als informatief beschouwd. Het toevoegen van deze factor heeft een zeer significante invloed op de dynamiek van de koersen van dealers. Meer specifiek zien wij dat alle dealers, met uitzondering van het handelssysteem Island, de neiging hebben om hun koersen van de beste koersen te verwijderen als zij deze bereikt hebben. Daarbij willen alle dealers hun prijzen dichter bij de beste prijs te zetten als het handelssysteem Island de beste prijzen heeft. Dit systeem wordt door veel daghandelaren gebruikt en aangezien op dit systeem anoniem gehandeld kan worden, wordt het systeem als zeer informatief beschouwd. Aanvullend bekijken we de efficiëntie van dealers met zogenaamde impulse response functies. Bij deze functies wordt een initiële schok aan het model gegeven waarna de impact van deze schok op langere termijn bekeken kan worden. Aangezien het model niet-lineaire eigenschappen bevat, moeten we een gegeneraliseerde vorm van deze functies toepassen. Deze gegeneraliseerde functies zijn gevoelig voor zowel de geschiedenis van het model, alswel de grootte van de schok die wordt gebruikt (een lineair model is ongevoelig voor deze feiten). Wij vinden dat de geschiedenis van het model de uitkomst van deze functies niet beïnvloedt. Echter het wel of niet toevoegen van deze niet-lineaire termen heeft een duidelijke invloed op de vorm van deze functies. De resultaten die wij met deze functies behalen, benadrukken nogmaals het belang van het handelssysteem Island.

In het vierde hoofdstuk stellen wij een nieuwe methode voor waarmee price discovery (price discovery is het process waarbij gekeken wordt in welke mate verschillende informatiebronnen invloed hebben op de prijsvorming van een aandeel) bekeken kan worden. De traditionele maatstaf voor price discovery is gedefinieerd als de hoeveelheid variantie die iedere beurs of, in dit geval, iedere dealer bijdraagt aan de variantie van de efficiënte prijs. Deze efficiënte prijs wordt in feite niet geobserveerd, maar is een geaccepteerde veronderstelling over het onderliggende prijsproces. Echter deze traditionele aanpak heeft de eigenschap dat het de mate van price discovery van iedere dealer niet uniek kan identificeren. Dit wordt hoofdzakelijk veroorzaakt door het feit dat data geaggregeerd wordt bij deze aanpak, met als gevolg dat informatie in de koers van een specifieke dealer al verwerkt is in de koersen van andere dealers (andere dealers hebben al de tijd gehad om te reageren). Indien dit het geval is kunnen we dus niet meer zeggen welke dealer de prijs heeft beïnvloedt. Het model dat wij

voorstellen is ontworpen voor data die op de hoogst mogelijke frequentie verkregen worden. De tijdschaal waarin deze data verworven worden, wordt ook wel tick tijd genoemd, waarbij iedere tick een verandering in koersen voorstelt. De werkelijke tijd (ook wel kalendertijd genaamd) die tussen de observaties in zit varieert in deze. Het ontworpen model beschouwt dus observaties in tick tijd, waarbij ook price discovery in tick tijd beschouwd wordt. In termen van price discovery definiëren wij drie verschillende maatstaven. De eerste geeft de reacties van dealers weer op veranderingen in de efficiënte prijs. Een tweede beschrijft de verandering van de efficiënte prijs ten gevolge van veranderingen in de koersen van dealers. Een derde maatstaf beschouwt de hoeveelheid variantie die iedere dealer bijdraagt aan de variantie van de efficiënte prijs. Wij tonen aan dat deze laatste maatstaf een combinatie is van de andere twee maatstaven. Verder zijn al deze maatstaven een functie van de tijd tussen koerswijzigingen. In het algemeen vinden wij dat koersen informatiever zijn en dat price discovery hoger is, als de tijd tussen koerswijzigingen kort is. Verder vinden we dat voor de drie verschillende maatstaven, verschillende dealers domineren. Als deze maatstaven voor price discovery over tijd geaggregeerd worden, convergeren deze naar een enkele maatstaf voor price discovery, welke vergelijkbaar is met de traditionele maatstaf. Echter, in dit geval leidt data aggregatie ertoe dat deze maatstaf naar een punt convergeert.

In hoofdstuk vijf beschouwen wij gerealiseerde volatiliteiten. Gerealiseerde volatiliteiten worden berekend door gekwadrateerde intra-dag rendementen over de dag te sommeren. Dit geeft een consistente maatstaf geeft voor de dagelijkse volatiliteit van een aandeel. De kracht van deze gerealiseerde volatiliteiten ligt in het feit dat deze geobserveerd zijn, in tegenstelling tot volatiliteitsschattingen die uit b.v. GARCH processen volgen en latent zijn. Echter als deze intra-dag rendementen op een zeer hoge frequentie gemeten worden, dan zijn deze besmet met ruis dat voortkomt uit allerlei microstructuur effecten (fricties en dergelijke). Dit leidt ertoe dat de gerealiseerde volatiliteit een inconsistente schatter is voor de werkelijke volatiliteit. In hoofdstuk vijf passen wij verschillende correcties toe op hoog frequente intra-dag rendementen. Hierbij maken we gebruik van twee verschillende tijdsschalen waarover data geaggregeerd kunnen worden, de normale kalendertijdschaal en een alternatieve transactietijdschaal. Ten eerste passen we een structureel model toe op transactiedata. Het toepassen van dit model verwijdert de negatieve autocorrelatie in rendementen, veroorzaakt door de fricties die invloed hebben op de microstructuur van de markt. Hierbij vinden we een ondergrens voor de gerealiseerde volatiliteit. Als tweede passen we een recent geïntroduceerde correctie toe. Wanneer deze correctie op transactiedata wordt toegepast, blijkt dat er een lange persistentie is in intra-dag transactierendementen. Deze persistentie wordt niet geobserveerd in kalendertijds-rendementen. Een correctie op deze persistentie leidt ertoe dat gerealiseerde volatiliteit gelijk is aan de volatiliteit die op lagere frequenties wordt gemeten. Het blijft echter een vraagstuk of deze persistentie een vorm van ruis is of dat deze inherent is aan het prijsproces.



Ten slotte beschouwt het laatste hoofdstuk de informatieve waarde van het transactieproces en het prijszettingproces. Het is een algeheel aanvaard begrip dat dealers toegang hebben tot publiekelijke informatie en op basis hiervan hun koersen zetten. Investeerders, daarentegen, kunnen toegang hebben tot private informatie. De informatie die in additie tot het prijszettingproces de markt instroomt via het transactieproces, is dan een maatstaf voor de hoeveelheid private informatie in de markt. De twee informatiecomponenten, publieke en private informatie, kunnen uit de transacties en de koersen geëxtraheerd worden door deze twee processen met elkaar te vergelijken. De motivatie hiertoe is dat beide processen hetzelfde onderliggende prijsproces volgen. Echter, transacties kunnen alleen plaatsvinden als koersen zijn gezet. Om te bepalen welke informatie (publiek of privé) het prijsproces drijft, laten we de innovatie van het prijsproces afhangen van het al dan niet plaatsvinden van een transactie of een koerswijziging. Over het algemeen blijkt dat transacties (private informatie) ongeveer 10% van de dagelijkse prijsverandering kunnen verklaren. Additionele toetsen op de private en publieke informatie onthullen dat er een zwakke relatie bestaat tussen private informatie en het verhandelde volume. Een sterke relatie bestaat er tussen het door dealers geplaatste volume en de publieke informatie. Echter een nog sterkere relatie vinden we tussen het aantal dealers dat gelijktijdig de beste koers plaatst en de publieke informatie. Verder blijkt dat de hoeveelheid publieke informatie groter is als een traditionele handelaar de beste koersen zet dan wanneer een alternatief handelssysteem de beste prijzen zet.

# Curriculum Vitae

Bart Frijns was born on April 4<sup>th</sup>, 1976 in Heerlen, the Netherlands. He studied Business Economics, with a first year in Econometrics at Maastricht University, the Netherlands. In 1999 he obtained his Master's Degree. After that he joined the department of finance at Maastricht University as a junior lecturer. One year later he became a PhD student at the same department, where he completed his PhD on empirical microstructure and volatility issues using (ultra) high frequency data in December 2004. His work has been presented at various international conferences around the world. From October 2004 he has been employed as a senior lecturer at Auckland University of Technology, New Zealand.

